

Query Match	11.6%	Score 408.8	DB 7	Length 2571
Best Local Similarity	60.0%	Pred: No. 1.8e-65		
Matches 823	Conservative 0	Mismatches 387	Indels 162	Gaps 2
1606	ACGAGAAATTTCTGATGGGATTAATCGAGGCTCATACCTGCATAGATTTTCACTATTTCCA 1665			
1185	AGCATATTTGTGCTACGAGAAATACAGATTTTCCAAATTCGACAGGTTTTCATTTTTTCTG 1244			
1666	TTCTGATCTAGAGAAACTATTTCTATGCCCCCGGATTTCCATCTCATATGATATCACTACAGTCA 1725			
1245	TTCAGATCTGATGAACCCGTGCTATGCTGATATATTCGCTCTCGTGTGTGATGCGAGGTCA 1304			
1726	ATCTGTCGAGATTTATTTTAAATATGSGTGAAGAAJACTGTGTGTGTAGATTTGATCTTGTC 1785			
1305	GTCAATTCATATGAATTTGTCTCAGAGGGGGGGAACCACTTGTGTGTGTAGATTTGCATTTGCC 1364			
1786	TACTGATTCGGAAATGAAAAATGTTGGCCAAAGCAATTTGGTATTCATCTCTTAACTGCTGA 1845			
1365	AACCATGACGAAATGCGTGTGCATATGCAAAAGGCTTTTGGAAATTCATCCATATGACGGCAGA 1424			
1846	AGATATTGCAATGCAAGAAJACTCGTGA AAAAGTGTGAATTTATTTAAAAGTTATTTATTTGT 1905			
1425	AGATATTAGAAATGCAAGAAJACTCGTGA AAAAGTGTGAGCGTTTTTAACTGCTGATTTATTTGT 1484			
1906	TTGTTTCCATACACTTTTGAAGCTGATAAAGAACTCTCAAGATATTTTAGAACCGATAAATGT 1965			
1485	TTGTTTCCATACATTTGAAAAAGCAAAAGAAATCGGAAGATTTCTTTAGAGCCCAATTAACGT 1544			
1966	TTATATTGTGTGTTTTTCCAATGATGATATTTAACTGTTTCCATTTTTCACCAATTTCTTCATCC 2025			
1545	TTACATGTCTGTTTGTAGATCTGGTGTGTTTGACCTTTCATTTTGGTTCCAATTTTCACATG 1604			
2026	AGCAAAATGTTTAGAAGAGAGTCTGCTCAATTTGAGAGATTTATGTCGATGTTAGTCTGATG 2085			
1605	TGCTAAATGTTAGAGGCTGTAAAGCACTACAGATTTAGCTCAACGCTCAACCTGATG 1664			

Result No.	Score	Query Match	Length	DB ID	Description
1	408.8	11.6	5571	7	US-11-217-529-581
2	403.8	11.5	5574	7	US-11-217-529-581
3	396	11.2	5580	7	US-11-217-529-78462
4	383	10.9	5577	7	US-11-217-529-78462
5	327.4	3.6	5910	7	US-11-217-529-77545
6	116.2	3.3	2910	7	US-11-217-529-3077
7	82.2	2.3	1134	7	US-11-217-529-166160
8	79.8	2.3	9859	6	US-10-517-441-456
9	79.8	2.3	9859	6	US-10-517-441-730
10	78.2	2.2	8759	6	US-10-517-441-256
11	78.2	2.2	8759	6	US-10-517-441-530
12	70.2	2.0	5286	6	US-10-517-441-294
13	70.2	2.0	5286	6	US-10-517-441-568
14	65.6	1.9	19634	6	US-10-517-441-688
15	63.6	1.8	2358	6	US-10-517-441-370
16	63.6	1.8	2358	6	US-10-517-441-644
17	63.6	1.8	13355	6	US-10-517-441-415
18	63.6	1.8	13355	6	US-10-517-441-689
19	62.6	1.8	9432	6	US-10-517-441-605
20	62.4	1.8	12610	6	US-10-517-441-441
21	62.4	1.8	12610	6	US-10-517-441-715
22	62.2	1.8	6343	6	US-10-517-441-686
23	61.4	1.7	10865	6	US-10-517-441-464
24	61.4	1.7	10865	6	US-10-517-441-738
25	61	1.7	5286	6	US-10-517-441-294
26	61	1.7	5286	6	US-10-517-441-568
27	61	1.7	19634	6	US-10-517-441-687
28	60.8	1.7	19634	6	US-10-517-441-414
29	60	1.7	13610	6	US-10-517-441-442
30	60	1.7	13610	6	US-10-517-441-716
31	59.4	1.7	4001	6	US-10-517-441-501
					Sequence 581, App
					Sequence 2281, App
					Sequence 78462, A
					Sequence 78462, A
					Sequence 77545, A
					Sequence 3077, App
					Sequence 166160,
					Sequence 456, App
					Sequence 730, App
					Sequence 256, App
					Sequence 530, App
					Sequence 294, App
					Sequence 568, App
					Sequence 688, App
					Sequence 370, App
					Sequence 644, App
					Sequence 415, App
					Sequence 689, App
					Sequence 605, App
					Sequence 441, App
					Sequence 715, App
					Sequence 686, App
					Sequence 464, App
					Sequence 738, App
					Sequence 294, App
					Sequence 568, App
					Sequence 687, App
					Sequence 414, App
					Sequence 442, App
					Sequence 716, App
					Sequence 501, App

[illegible]

Sequence 23, Appl
Sequence 27, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 1038, Ap
Sequence 14164, A
Sequence 12378, A
Sequence 26, Appl
Sequence 12776, A
Sequence 15940, A
Sequence 25, Appl
Sequence 16536, A
Sequence 20, Appl

30 65.6 1.9 50000 3 US-09-662-254B-23
31 65.4 1.9 32392 3 US-09-662-254B-27
32 65.2 1.9 1037 3 US-09-181-585-3
33 65.2 1.9 1159 3 US-09-181-585-1
34 65.2 1.9 1471 3 US-09-181-585-2
35 65 1.8 700 3 US-09-735-271-1038
36 65 1.8 18773 3 US-09-949-016-14164
37 64 1.8 1141 3 US-09-806-708B-22
38 63.6 1.8 50000 3 US-09-662-254B-26
39 63.6 1.8 119153 3 US-09-949-016-12378
40 63.4 1.8 837 3 US-08-998-416-288
41 63.4 1.8 187169 3 US-09-949-016-12776
42 63.4 1.8 191569 3 US-09-949-016-15940
43 63.2 1.8 50000 3 US-09-662-254B-25
44 62.4 1.8 114139 3 US-09-949-016-16536
45 62 1.8 4441 3 US-09-796-348-20

ALIGNMENTS

RESULT 1
US-09-248-796A-6328
; Sequence 6328, Application US/09248796A
; Patent No. 674737
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA
ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 6328
; LENGTH: 2358
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-6328

Query Match 61.28; Score 2155.2; DB 3; Length 2358;
Best Local Similarity 97.28; Pred. No. 0;
Matches 2293; Conservative 0; Mismatches 23; Indels 46; Gaps 8;

QY 285 ATAAATATAATTAAGAGTATATCTCCCTCTTTTGTGTTTTTCTTCCAGCCATGTCG 344
DB 1 ATAAATATAATTAAGAGTATATCTCCCTCTTTTGTGTTTTTCTTCCAGCCATGTCG 58
QY 345 ATAGTGAAGTTTATTAATCAAAATTCATCTAATCAACCTATCTCTAGATTCATGAAG 404
DB 59 ATAGTGAAGTTTATTAATCAAAATTCATCTAATCAACCTATCTCTAGATTCATGAAG 118
QY 405 TATTGGATGATCATAGAAATCAAAATCACTAATGATTTGTCATAGTAGTGAAGATG 464
DB 119 TATTGGATGATCATAGAAATCAAAATCACTAATGATTTGTCATAGTAGTGAAGATG 178
QY 465 AGTTGGAAATTAATAATTCAGATTAAGATTCAGAAAGTTGTAAAGAGCAAA----- 512
DB 179 AGTTGGAAATTAATAATTCAGATTAAGATTCAGAAAGTTGTAAAGAGCAAA----- 238
QY 513 AACCAACACATCATCAAGAGATTCATCAAGATTAATGCTAAACCATTCCTGTAAT 572
DB 239 AACCAACACATCATCAAGAGATTCATCAAGATTAATGCTAAACCATTCCTGTAAT 298
QY 573 CTGGTTCTTCAATTAAGAAAATCTTAATCTTACCGATAAGATAGATTTACCAACCTTA 632
DB 299 CTGGTTCTTCAATTAAGAAAATCTTAATCTTACCGATAAGATAGATTTACCAACCTTA 358
QY 633 TCAGTTATCTGGTGGTATGATTAATAACACCGGTCACAAAATTCGTAAATTAACA 692
DB 359 TCAGTTATCTGGTGGTATGATTAATAACACCGGTCACAAAATTCGTAAATTAACA 418
QY 693 TCAGTTATCTGGTGGTATGATTAATAACACCGGTCACAAAATTCGTAAATTAACA 752
DB 419 TCAGTTATCTGGTGGTATGATTAATAACACCGGTCACAAAATTCGTAAATTAACA 478

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.
OH nucleic - nucleic search, using sw model
Run on: June 15, 2006, 03:11:16 ; Search time 653 Seconds
(without alignments)
10100.549 Million cell updates/sec

Title: US-10-018-105A-3
Perfect score: 3523
Sequence: 1 tatataataataataata.....tcaattattatcaatattgc 3525
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues
Total number of hits satisfying chosen parameters: 2807332
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /EMC_Celerra_SIDS3/ptodata/2/ina/1_COMB.seq.*
2: /EMC_Celerra_SIDS3/ptodata/2/ina/5_COMB.seq.*
3: /EMC_Celerra_SIDS3/ptodata/2/ina/6A_COMB.seq.*
4: /EMC_Celerra_SIDS3/ptodata/2/ina/6B_COMB.seq.*
5: /EMC_Celerra_SIDS3/ptodata/2/ina/7_COMB.seq.*
6: /EMC_Celerra_SIDS3/ptodata/2/ina/7_COMB.seq.*
7: /EMC_Celerra_SIDS3/ptodata/2/ina/H_COMB.seq.*
8: /EMC_Celerra_SIDS3/ptodata/2/ina/PTUS_COMB.seq.*
9: /EMC_Celerra_SIDS3/ptodata/2/ina/PP_COMB.seq.*
10: /EMC_Celerra_SIDS3/ptodata/2/ina/RE_COMB.seq.*
10: /EMC_Celerra_SIDS3/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2155.2	61.2	2358	3	US-09-248-796A-6328
2	423.8	12.0	507	3	US-09-248-796A-6327
3	168.2	4.8	537	3	US-08-998-416-66
4	118.2	3.4	2445	3	US-09-248-796A-6302
5	116.4	3.3	688	3	US-09-533-559-7103
6	106.2	3.0	575	3	US-09-533-559-447
7	74.4	2.1	50000	3	US-09-662-254B-23
8	73.4	2.1	708	3	US-09-248-796A-10622
9	71.8	2.0	19124	2	US-08-487-826B-13
10	70.4	2.0	50000	3	US-09-662-254B-25
11	70.2	2.0	18651	3	US-09-949-002-592
12	70.2	2.0	18682	3	US-09-949-002-786
13	68.8	2.0	1141	3	US-08-906-708B-22
14	68.2	1.9	50000	3	US-09-662-254B-24
15	68	1.9	767677	3	US-09-949-016-12147
16	67.6	1.9	767677	3	US-09-949-016-17361
17	67.6	1.9	29717	3	US-09-949-016-16284
18	67.6	1.9	32392	3	US-09-662-254B-27
19	67.6	1.9	60376	3	US-09-949-016-12423
20	66.8	1.9	1410	3	US-08-601-198-56
21	66.8	1.9	14066	3	US-09-662-254B-33
22	66.4	1.9	19438	3	US-09-949-016-12699
23	66	1.9	26000	3	US-08-843-376-10
24	65.8	1.9	205044	3	US-09-949-016-15851
25	65.8	1.9	205044	3	US-09-949-016-15852
26	65.8	1.9	205044	3	US-09-949-016-15853
27	65.8	1.9	223471	3	US-09-949-016-12387
28	65.8	1.9	223471	3	US-09-949-016-12724
29	65.8	1.9	223471	3	US-09-949-016-12725

[illegible]

32 113 2.3 877 6 US-10-471-571A-3884 Sequence 3884, Ap
33 111.5 2.3 560 6 US-10-953-349-3923 Sequence 3923, Ap
34 111.5 2.3 1349 6 US-10-471-571A-3352 Sequence 3352, Ap
35 111 2.3 340 6 US-10-953-349-20922 Sequence 20922, A
36 111 2.3 613 7 US-11-293-697-4125 Sequence 4125, A
37 111 2.3 619 6 US-10-471-571A-3914 Sequence 3914, Ap
38 111 2.3 895 6 US-10-471-571A-922 Sequence 922, App1
39 111 2.3 2829 6 US-10-525-621-1 Sequence 1, App1
40 110.5 2.3 568 6 US-10-953-349-10556 Sequence 10556, A
41 110.5 2.3 718 6 US-10-953-349-5389 Sequence 5389, Ap
42 110.5 2.3 738 6 US-10-953-349-5388 Sequence 5388, Ap
43 110.5 2.3 763 6 US-10-953-349-5387 Sequence 5387, Ap
44 110.5 2.3 1073 7 US-11-293-697-2915 Sequence 2915, Ap
45 110.5 2.3 1809 6 US-10-370-959-67 Sequence 67, App1

ALIGNMENTS

RESULT 1
US-10-471-571A-3948
; Sequence 3948, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471.571A
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 3948
; LENGTH: 2271
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(2271)
; OTHER INFORMATION: hypothetical protein
US-10-471-571A-3948

Query Match 3.5%; Score 166.5; DB 6; Length 2271;
Best Local Similarity 18.3%; Pred. No. 0.052;
Matches 109; Conservative 103; Mismatches 234; Indels 151; Gaps 20;

QY 1 MDSSEYQNSTNQIPRSDEVDHNRQITNDCASDSEDELEKSELESEVVKSEKQ 60
DB 1701 ISDSQSMSESVNDSSESVSESSDS--KMSGTSVSDS--GSLSVSTSLRKSESVSESS 1757
QY 61 QQRHQETSDNAKPIVTRKS---GSSIKKSNLTQKDRITWPKSLGGDDTIHSGHKRNY 117
DB 1758 SLSCSQMSDSVSTSDSSSVSTSLASSESVSESDSLSDSKSTSGSTSTSGSLSTP 1817
QY 118 NMS---SLRKDFYKDN---TQDNSTNNHTLAIPPIPIPTPIIT-----NANKSR 163
DB 1818 SLGSESVSESTSLSDSISMSDSTSTSDSLSGSISLSGSLSTSLSDSLSDSKSLSSQ 1877
QY 164 RKSQENLPLIKKTKTGRNNNNFENDLV--SPMTKTKTNDSEDTITTTTANRKLIG- 220
DB 1878 SMSGSESTSVSDSQSSSTSVSDSQSSSTSVSDSQSSSTSVSDSQSSSTSVSDSQSS 1937
QY 221 -----IGATLTGVTGTAT-----ATATAAGR----- 244
DB 1938 SMSGSESVSTSLSDSISGTSVSDSSTSTSTSLSDSMSQSOSTSTACGSLSTST 1997
QY 245 -----RPSRSSIDSEADS-----HARRSQETEDVCFPMVGDHVRNG 283
DB 1998 SMSMASTSSQSTSVSTSLSDSISGTSVSDSSTSTSLSDSISGTSVSESTSDSISDS 2057
QY 284 IDPDIDEFIREREAYLQKQAKNLAIRIDEFQNLKNNITGARRHPYHHNNKK 343
DB 2058 SDSDS-----TSTSTSDSTSGT-----STSISESL 2083
QY 344 NNGGSGGSMALKYTPKNILKLTLSRPFTHENSSESEIYELKTKQPPKYDDQLS 403
DB 2084 STSGSGSTSVSDSTSMSESVSSVSDSQSDSTSDSISDSSESV---STSTSTSLSDSTS 2140

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: June 13, 2006, 14:21:02 ; Search time 17 Seconds
(without alignments)
685.708 Million cell updates/sec
Title: US-10-018-105A-4
Perfect score: 4820
Sequence: 1 MDSSEYQNSTNQIPRS.....LGLKKGNGKNSIIFPNKY 922

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 64916 seqs, 12643201 residues

Total number of hits satisfying chosen parameters: 64916

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New:

1: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US09_NEW_PUB pep:
2: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US06_NEW_PUB pep:
3: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US07_NEW_PUB pep:
4: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US08_NEW_PUB pep:
5: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/PCT_NEW_PUB pep:
6: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US10_NEW_PUB pep:
7: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US11_NEW_PUB pep:
8: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US60_NEW_PUB pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	166.5	3.5	2271	6	US-10-471-571A-3948 Sequence 3948, Ap
2	157	3.3	750	7	US-11-134-228A-32 Sequence 32, App1
3	137	2.8	454	7	US-11-134-228A-35 Sequence 35, App1
4	137	2.8	1392	6	US-10-953-349-1430 Sequence 1430, Ap
5	136	2.8	315	6	US-10-471-571A-4384 Sequence 4384, Ap
6	136	2.8	2478	6	US-10-471-571A-2278 Sequence 2278, Ap
7	135.5	2.8	596	6	US-10-196-749-310 Sequence 310, App
8	135.5	2.8	596	7	US-11-101-316-100 Sequence 100, App
9	133	2.8	933	7	US-11-121-154-178 Sequence 178, App
10	130	2.7	1274	6	US-10-471-571A-898 Sequence 898, App1
11	129.5	2.7	624	6	US-10-528-104-3 Sequence 3, App1
12	128.5	2.7	488	6	US-10-471-571A-10 Sequence 10, App1
13	128.5	2.7	788	7	US-11-293-697-4150 Sequence 4150, App
14	126	2.6	710	7	US-11-134-228A-33 Sequence 33, App1
15	124.5	2.6	477	7	US-11-134-228A-34 Sequence 34, App1
16	124.5	2.6	935	6	US-10-471-571A-4496 Sequence 4496, Ap
17	123	2.6	688	7	US-11-293-697-2798 Sequence 2798, App
18	121	2.5	940	7	US-11-293-697-3781 Sequence 3781, Ap
19	119.5	2.5	774	6	US-10-471-571A-4824 Sequence 4824, Ap
20	119.5	2.5	791	7	US-11-121-154-24 Sequence 24, App1
21	119.5	2.5	1106	7	US-11-293-697-4301 Sequence 4301, App
22	119	2.5	502	6	US-10-511-937-3008 Sequence 3008, Ap
23	118	2.4	387	6	US-10-953-349-8568 Sequence 8568, Ap
24	118	2.4	439	6	US-10-953-349-8567 Sequence 8567, Ap
25	117	2.4	1003	6	US-10-953-349-1431 Sequence 1431, Ap
26	116.5	2.4	2937	6	US-10-480-962-18 Sequence 18, App1
27	116.5	2.4	2969	6	US-10-480-962-19 Sequence 19, App1
28	116	2.4	1075	6	US-11-134-228A-23 Sequence 23, App1
29	115.5	2.4	1261	6	US-10-471-571A-1304 Sequence 1304, Ap
30	113	2.3	611	6	US-10-953-349-3922 Sequence 3922, Ap
31	113	2.3	624	6	US-10-953-349-3921 Sequence 3921, Ap

OM protein - protein search, using sw model

Run on: June 13, 2006, 14:08:32 ; Search time 52 Seconds
(without alignments)
1551.985 Million cell updates/sec

Title: US-10-018-105A-4

Perfect score: 4920

Sequence: 1 MSDSYQSTNQIPRSD.....LGLKHGNGKSIISPPNKE 922

Scoring table: BLOSUM62

Gap 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_Aa.*

- 1: /EMC_Celerra_SID33/ptodata/2/1aa/5_COMB.pep.*
- 2: /EMC_Celerra_SID33/ptodata/2/1aa/6_COMB.pep.*
- 3: /EMC_Celerra_SID33/ptodata/2/1aa/7_COMB.pep.*
- 4: /EMC_Celerra_SID33/ptodata/2/1aa/8_COMB.pep.*
- 5: /EMC_Celerra_SID33/ptodata/2/1aa/9_COMB.pep.*
- 6: /EMC_Celerra_SID33/ptodata/2/1aa/RE_COMB.pep.*
- 7: /EMC_Celerra_SID33/ptodata/2/1aa/backfilled.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3785	78.5	2	US-09-248-796A-20431	Sequence 20431, A
2	1880	34.3	859	US-09-538-092-1371	Sequence 717, App
3	1734	17.3	814	US-09-248-796A-20405	Sequence 20405, A
4	716	14.3	188	US-09-248-796A-20430	Sequence 20430, A
5	251	5.2	326	US-09-502-540-11640	Sequence 11640, A
6	172.5	3.6	1051	US-09-538-092-469	Sequence 469, App
7	166.5	3.5	2283	US-10-172-502-4	Sequence 4, App1
8	166	3.4	719	US-09-248-796A-17559	Sequence 17559, A
9	161	3.3	366	US-09-712-363-193	Sequence 193, App
10	161	3.3	471	US-08-311-731A-168	Sequence 168, App
11	159.5	3.3	1177	US-09-134-001C-5106	Sequence 5106, App
12	159.5	3.3	1253	US-09-958-617A-18	Sequence 18, App1
13	159.5	3.3	1584	US-09-457-040B-27	Sequence 27, App1
14	159.5	3.3	2150	US-10-135-322-17	Sequence 17, App1
15	158.5	3.3	785	US-09-487-558B-348	Sequence 348, App
16	158	3.3	651	US-08-431-080-24	Sequence 24, App1
17	158	3.3	651	US-08-938-534-24	Sequence 24, App1
18	158	3.3	651	US-09-345-294-24	Sequence 24, App1
19	157	3.3	858	US-09-248-796A-19055	Sequence 19055, A
20	154	3.2	1259	US-09-949-016-10366	Sequence 10366, A
21	153.5	3.2	982	US-09-248-796A-20628	Sequence 20628, A
22	152	3.2	2870	US-09-479-467A-15	Sequence 15, App1
23	152	3.2	2870	US-09-655-160-15	Sequence 15, App1
24	152	3.2	3178	US-09-479-467A-4	Sequence 4, App1
25	152	3.2	3178	US-09-655-160-4	Sequence 4, App1
26	151	3.1	315	US-09-710-279-2100	Sequence 2100, App
27	151	3.1	324	US-09-134-001C-4080	Sequence 4080, App
28	151	3.1	10182	US-09-134-001C-3159	Sequence 3159, App
29	150	3.1	676	US-09-134-001C-4318	Sequence 4318, App
30	147	3.0	404	US-09-710-279-2964	Sequence 2964, App
31	147	3.0	5024	US-09-710-279-2964	Sequence 2964, App
32	146.5	3.0	1444	US-09-949-016-9652	Sequence 9652, App

33	146.5	3.0	1667	US-09-270-767-41425	Sequence 41425, A
34	146.5	3.0	2004	US-09-538-092-1371	Sequence 1371, App
35	146.5	3.0	2004	US-09-949-016-6756	Sequence 6756, App
36	146	3.0	292	US-09-248-796A-25373	Sequence 25373, A
37	145	3.0	873	US-09-248-796A-20365	Sequence 20365, A
38	145	3.0	1435	US-08-568-459A-4	Sequence 4, App1
39	145	3.0	1435	US-08-487-826B-4	Sequence 4, App1
40	145	3.0	1435	US-09-210-288-4	Sequence 4, App1
41	145	3.0	1435	US-10-153-273-4	Sequence 4, App1
42	144.5	3.0	1085	US-08-431-080-28	Sequence 28, App1
43	144.5	3.0	1085	US-08-938-534-28	Sequence 28, App1
44	144.5	3.0	1085	US-09-345-294-28	Sequence 28, App1
45	144.5	3.0	3092	US-09-487-558B-172	Sequence 172, App

ALIGNMENTS

RESULT 1

US-09-248-796A-20431
; Sequence 20431, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/7074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20431
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20431

Query Match 78.5%; Score 3785; DB 2; Length 785;
Best Local Similarity 96.2%; Pred. No. 3e-296;
Matches 740; Conservative 2; Mismatches 13; Indels 14; Gaps 7;

QY	1	MSDSYQSTNQIPRSDVLDHRRNQITNDCAISDSDELEKSELESEVAKSEK	- 59
DB	18	MSDSYQSTNQIPRSDVLDHRRNQITNDCAISDSDELEKSELESEVAKSEK	77
QY	60	---QQQHHEITSDNAKPLTKSGSISIKKSNLTDKRITNPMSLSGGDDTINSCHKRN	116
DB	78	QQQQQQHHEITSDNAKPLTKSGSISIKKSNLTDKRITNPMSLSGGDDTINSCHKRN	137
QY	117	YNMSSLRKDFYKXNDTNNHNLHAIPIPIPTIITNANKSRKSOLENLPLIK	176
DB	138	YNMSSLRKDFYKXNDTNNHNLHAIPIPIPTIITNANKSRKSOLENLPLIK	197
QY	177	KKTIGRNNNNPNDLVSPMTQKTNDSIEDITN-TSTTANMKLGIGATTGLVGTTAT	235
DB	198	KKTIGRNNNNPNDLVSPMTQKTNDSIEDITN-TSTTANMKLGIGATTGLVGTTAT	257
QY	236	ATATATAA---GRPSRSSISSEADSHASRSQSQTEDVCFPMVGDHVRVNGIDFDEIDEP	292
DB	258	ATATATAGGGRPSRSSISSEADSHASRSQSQTEDVCFPMVGDHVRVNGIDFDEIDEP	317
QY	293	IREEREAYLQKMAKNIILRIDEFQNTSKNNTTSGASRPHYHHNNKKNNGDG-GG	351
DB	318	IREEREAYLQKMAKNIILRIDEFQNTSKNNTTSGASRPHYHHNNKKNNGDGSGG	377
QY	352	SSMAALKYTPNKLKTLSPFETHNSSSSEIYELKTKQPPKYDDQLSLTSTSTST	411
DB	378	SSMAALKYTPNKLKTLSPFETHNSSSSEIYELKTKQPPKYDDQLSLTSTSTST	437
QY	412	--SGSGSGQVKGARISDINGSLDPFSLPHSESETHAPDIPSLVSPQSVROLF	469
DB	438	SGSGSGQVKGARISDINGSLDPFSLPHSESETHAPDIPSLVSPQSVROLF	497
QY	470	RNGEETWMLDCTCTDSEMDCLAKAFGHPHTAEDIRMOETREKVELFKSVYVCFHTE	529

34 165 3.4 4498 6 US-11-097-143-2577 Sequence 2577, Ap
35 164.5 3.4 2271 4 US-10-282-122A-43924 Sequence 43924, A
36 161 3.3 366 3 US-09-712-363-193 Sequence 193, App
37 161 3.3 2719 5 US-10-732-923-8668 Sequence 8668, Ap
38 160 3.3 373 2 US-08-945-749-9 Sequence 9, Appli
39 159.5 3.3 1169 4 US-10-282-122A-71179 Sequence 71179, A
40 159.5 3.3 1177 4 US-10-724-972A-6013 Sequence 6013, Ap
41 159.5 3.3 1253 4 US-10-363-928-2 Sequence 2, Appli
42 159.5 3.3 1253 6 US-11-185-924-18 Sequence 18, Appli
43 159.5 3.3 1584 5 US-10-732-923-13548 Sequence 13548, A
44 159.5 3.3 2150 4 US-10-135-322-17 Sequence 17, Appli
45 159 3.3 2399 5 US-10-732-923-15036 Sequence 15036, A

ALIGNMENTS

RESULT 1
US-10-032-585-7675
; Sequence 7675, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032.585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7675
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7675

Query Match 99.9%; Score 4814; DB 4; Length 922;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 921; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSDESYQNSTNQIPRSDVLDIRNQITNDCAISDESELEKSELESEVVKSEKQ 60
Db 1 MSDESYQNSTNQIPRSDVLDIRNQITNDCAISDESELEKSELESEVVKSEKQ 60
Qy 61 QOHQETSDNAKPLTRKSGSSIKKSNLTKDRITNPMSLSGDDTINSCHKNNYNS 120
Db 61 QOHQETSDNAKPLTRKSGSSIKKSNLTKDRITNPMSLSGDDTINSCHKNNYNS 120
Qy 121 SLRQDFYKDMTDDNNTHLAIPIPIPIPIITNANKSRKSKLENLPLKKKTI 180
Db 121 SLRQDFYKDMTDDNNTHLAIPIPIPIPIITNANKSRKSKLENLPLKKKTI 180
Qy 181 GRNNSNPNFNDLVSPMTKKTNDSEDITNTSTANHKMLGIGATTGVTGTTATATA 240
Db 181 GRNNSNPNFNDLVSPMTKKTNDSEDITNTSTANHKMLGIGATTGVTGTTATATA 240
Qy 241 AAGRPRSSIDSEADSHASRSQETEDVCPFMVGDIHVNGIDFDEIDFIREEAA 300
Db 241 AAGRPRSSIDSEADSHASRSQETEDVCPFMVGDIHVNGIDFDEIDFIREEAA 300
Qy 301 YLOQMIAKNILRIDEQNLSKANTTSGASRHPYHHNSNNKNGGSGGSSMAALKYT 360
Db 301 YLOQMIAKNILRIDEQNLSKANTTSGASRHPYHHNSNNKNGGSGGSSMAALKYT 360
Qy 361 PNILKLTLSRFEFTHENSSESEIYELTKQQPYKDDQLSLTSTSTSGSGSQVR 420
Db 361 PNILKLTLSRFEFTHENSSESEIYELTKQQPYKDDQLSLTSTSTSGSGSQVR 420
Qy 421 FGARISDGINSGSLPDRFSLPHSESETHAPDIPSLVPSQSVRLDFNGEETWMD 480
Db 421 FGARISDGINSGSLPDRFSLPHSESETHAPDIPSLVPSQSVRLDFNGEETWMD 480
Qy 481 TCFDSEKMLAKAFGHPHTAEDIRMQETREKVELPKSYVFCFHTFEADKESDYEP 540
Db 481 TCFDSEKMLAKAFGHPHTAEDIRMQETREKVELPKSYVFCFHTFEADKESDYEP 540
Qy 541 INVYIVFHGILTFHFSPISHPANVRVRQLRDYVDVSADMLCYALIDEITDGFAPVI 600

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: June 13, 2006, 14:20:38 ; Search time 182 Seconds
(without alignments)
2346.616 Million cell updates/sec

Title: US-10-018-105A-4
Perfect score: 4820
Sequence: 1 MSDESYQNSTNQIPRSDVLDIRNQITNDCAISDESELEKSELESEVVKSEKQ 922
Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5
Searched: 2097797 seqs, 463214858 residues
Total number of hits satisfying chosen parameters: 2097797
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:
1: /EMC_Celerra-SIDS3/ptodata/2/pubpa/US07_PUBCOMB.pcp:
2: /EMC_Celerra-SIDS3/ptodata/2/pubpa/US08_PUBCOMB.pcp:
3: /EMC_Celerra-SIDS3/ptodata/2/pubpa/US09_PUBCOMB.pcp:
4: /EMC_Celerra-SIDS3/ptodata/2/pubpa/US10A_PUBCOMB.pcp:
5: /EMC_Celerra-SIDS3/ptodata/2/pubpa/US10B_PUBCOMB.pcp:
6: /EMC_Celerra-SIDS3/ptodata/2/pubpa/US11_PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4814	99.9	922	4 US-10-032-585-7675	Sequence 7675, Ap
2	1679	34.8	859	2 US-08-945-749-1	Sequence 1, Appli
3	1666	34.6	858	2 US-08-945-749-2	Sequence 2, Appli
4	1099	22.8	663	4 US-10-128-714-3563	Sequence 3563, Ap
5	1099	22.8	663	4 US-10-128-714-8563	Sequence 8563, Ap
6	856	17.8	969	2 US-08-945-749-3	Sequence 3, Appli
7	345.5	7.2	154	3 US-09-864-408A-2266	Sequence 2266, Ap
8	298	6.2	94	4 US-10-425-115-321932	Sequence 321932, Ap
9	234.5	4.9	387	2 US-08-945-749-8	Sequence 8, Appli
10	234.5	4.9	387	2 US-11-097-143-7905	Sequence 7905, Ap
11	213.5	4.4	380	2 US-08-945-749-7	Sequence 7, Appli
12	213.5	4.4	386	4 US-10-156-761-9574	Sequence 9574, Ap
13	201	4.2	371	4 US-10-156-761-10644	Sequence 10644, Ap
14	182.5	3.8	373	3 US-09-738-626-3568	Sequence 3568, Ap
15	179.5	3.7	373	3 US-11-082-389-248	Sequence 248, App
16	179.5	3.7	1172	4 US-10-451-467A-574	Sequence 574, App
17	179.5	3.7	2344	4 US-10-732-923-13978	Sequence 13978, A
18	179	3.7	354	4 US-10-156-761-13791	Sequence 13791, A
19	179	3.7	1199	6 US-11-097-143-1614	Sequence 1614, Ap
20	175.5	3.6	1246	6 US-11-097-143-11433	Sequence 11433, A
21	172.5	3.6	3328	5 US-10-732-923-8311	Sequence 8311, Ap
22	172	3.6	2165	4 US-10-732-923-13547	Sequence 13547, A
23	170.5	3.5	343	4 US-10-156-761-14171	Sequence 14171, A
24	169	3.5	1468	6 US-11-097-143-22779	Sequence 22779, A
25	168.5	3.5	816	6 US-10-467-555-32	Sequence 32, Appli
26	166.5	3.5	1046	6 US-11-165-819-6	Sequence 6, Appli
27	166.5	3.5	1046	6 US-11-165-819-26	Sequence 26, Appli
28	166.5	3.5	2261	4 US-10-470-048B-60	Sequence 60, Appli
29	166.5	3.5	2283	4 US-10-172-503-4	Sequence 4, Appli
30	166.5	3.5	2283	4 US-11-020-509-4	Sequence 4, Appli
31	166	3.4	688	4 US-10-032-585-7876	Sequence 7876, Ap
32	165	3.4	2112	6 US-11-097-143-8001	Sequence 8001, Ap
33	165	3.4	4498	4 US-10-712-124-6081	Sequence 68, Appli

Db 541 INVYVWFHDLGTLTFHSPISHANVRVRQLRDYVDVSDMLCYALIDEITDGPAPI 600
Qy 601 HGIEYEADEADAVTARDTDFSNLQRIGESRRKVTMLRLLSGKADVIKMFARQBE 660
Db 601 HGIEYEADEADAVTARDTDFSNLQRIGESRRKVTMLRLLSGKADVIKMFARQBE 660
Qy 661 ANSSSGYQRYNLOOQOQAPPPPPNPIITSPINSLNLSLGTSTGGVGVGGINFGP 720
Db 661 ANSSSGYQRYNLOOQOQAPPPPPNPIITSPINSLNLSLGTSTGGVGVGGINFGP 720
Qy 721 NPTGNNTNTNTTSGSPSPPOOQOHHGINTKSPIDARPRADIALYLGDIQDHIITMQ 780
Db 721 NPTGNNTNTNTTSGSPSPPOOQOHHGINTKSPIDARPRADIALYLGDIQDHIITMQ 780
Qy 781 NLLAYEKFIRSHSNLYLAQLQVESFNSNNKITEMFSKITLIGTMLVPLNLVTLGFGNVR 840
Db 781 NLLAYEKFIRSHSNLYLAQLQVESFNSNNKITEMFSKITLIGTMLVPLNLVTLGFGNVR 840
Qy 841 VFBEGTNLQWFGVGLVPIIIGSFIFAQWMLKLNLSIEQNGNRPFIHSSRSI 900
Db 841 VFBEGTNLQWFGVGLVPIIIGSFIFAQWMLKLNLSIEQNGNRPFIHSSRSI 900
Qy 901 RSLGLKHGKNSIISPPNKE 922
Db 901 RSLGLKHGKNSIISPPNKE 922
RESULT 2
US-08-945-749-1
; Sequence 1, Application US/08945749
; Publication No. US2002013880A1
; GENERAL INFORMATION:
; APPLICANT: GARDNER, Richard C
; APPLICANT: MacDIARMID, Colin W
; APPLICANT: HAY, Robert J
; APPLICANT: Auckland Uniservices Limited
; APPLICANT: New Zealand Pastoral Agriculture Research Institut
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE
; FILE REFERENCE: 08/945,749
; CURRENT APPLICATION NUMBER: US/08/945,749
; EARLIER FILING DATE: 1998-01-12
; EARLIER FILING DATE: PCT/NZ96/0035
; EARLIER FILING DATE: 1996-05-01
; EARLIER FILING DATE: NZ 272039
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-08-945-749-1
Query Match 34.8%; Score 1679; DB 2; Length 859;
Best Local Similarity 41.5%; Pred. No. 3.5e-107;
Matches 405; Conservative 137; Mismatches 258; Indels 176; Gaps 31;
Qy 1 MDSSEYQNTNQPPIRSDVLDHRRNQINDCAISDSDELEL---KSESESVKS 57
Db 1 MSSSS---SSESPYLSRS-----NSLANTMYSMTEDHTGLYDHRQHPDLPVRH 49
Qy 58 EKQOHHQBITSDNAKPLTRKSGSIKKS-----NLTDKRIITNPMISLGGDDTINSGH 112
Db 50 QPTLKNKSIKASTKPSIPKQKATRYNSHVGVSPSRGMDPDEQCGHSDTV-AHH 108
Qy 113 KNR-----NVMSSI-----RKDFYKNDTDSNTNMT---HLAIPPIPIPTPII 156
Db 109 QURASALLTNSRPSRLAHSMHQRLYVESNI-----HTPPKDVGVKRDYTMSSSTA 161
Qy 157 TMANKSRKSOLELPLKKKTIIRNNNNFENDLVSPMTQKTNDSIEDITNT-----210
Db 162 SSGNKS-KLSASSASPIKTVR-----KSLVSPVLEIPHEKSDTHSLAKPKK 210
Qy 211 ---STTANMKLGIGATTGTVGCTTATA---TATAAAGRPSRSSISDSEADSHSRSSQ 264
Db 211 RYTYTSAHSIN-PAVLITKSTSQSDADDTLERKPYNMVMTASFSDV-SQGRDSQ 268
Qy 265 ETEEDVCFPMVGD-HIRVNGIDFDEIREREAYLQKQMIK-NILRIDFQNLK 322

Db 269 ETEEDVCFPMVGPQLTRVNGIDFDEIRERAYAFANAE---KSQFLASQVNPBOKYSNVSQ 325
Qy 323 -----NNTTSGARHPHYHHNNKNNKNGGGGSGMAALKYTPKILAKTLRSRFFTH 376
Db 326 DIGFTSTSTSGSS-----AALKYTPR--VSQTCBKSESTN 359
Qy 377 ENSSSSEIYELKTKOQPPKYDDQLSLTSSSTSGSGGVKFGG-ARISDGINGGSL 435
Db 360 ET-----EHEKKEDEHEKIKPSLHPGISFGKNKVEGEENENIPSNDAVCSYQGTDFQI 414
Qy 436 PDRSLFSHSESETHAPDIPSLVSPQSVROLFRNGEETWMLDCTPTDSEMGMKAKAF 495
Db 415 PNRSPFCSESEDETVHASDTPSLVSEGOFTFELFRGGEPTWMLDSCPTDDMRCIAKAF 474
Qy 496 GHPUATAEDIRMOETREKVELFKSYTFCFTFPAKSESESYLEPINVIYVWFHDLGTLT 555
Db 475 GHPUATAEDIRMOETREKVELFKSYTFCFTFPAKSESESYLEPINVIYVWFHDLGTLT 534
Qy 556 HFSPIHPANVRVRQLRDYVDVSDMLCYALIDEITDGPAPVHIEYEADAEADAVP 615
Db 535 HFGPISHCANVRVRQLRDYVDVSDMLCYALIDEITDGPAPVHIEYEADAEADAVP 594
Qy 616 TARDTDFSNLQRIGESRRKVTMLRLLSGKADVIKMFARQBEANSSSGYQRYNLOQ 675
Db 595 MARDMDFAMLQRIGESRRKVTMLRLLSGKADVIKMFARQBEANGI-----643
Qy 676 OOOQOAPPPPPNPIITSPINSLNLSLGTSTGGVGVGGINFGNPTGNNTNTNTTGT 735
Db 644 -----GPAULTSQIN-IANLQARQDNA-----SHIKNNSSTVTPNNY 678
Qy 736 SPSPPOOQOHHGINTKSPIDARPRADIALYLGDIQDHIITMQLNLAVERIFSRSHN 795
Db 679 APTTSQ-----PRGDIALYLGDIQDHIITMQLNLAVERIFSRSHN 720
Qy 796 YLAQLQVESFNSNNKITEMFSKITLIGTMLVPLNLVTLGFGNVRVFBEGTNLQWFGFI 855
Db 721 YLAQLQVESFNSNNKITEMFSKITLIGTMLVPLNLVTLGFGNVRVFBEGTNLQWFGFI 779
Qy 856 VGLVPIIIGSFIFAQWMLKLNLSI---EQNGNRPFIHSSRSIRSLGK 907
Db 780 LGVLLLVGLVFLASVYKIRIDPPATLNEAESKASVSISSFLPKRKNRPNDRSKNIN 839
Qy 908 HGG--NKSIISPPNKE 921
Db 840 RAGFSNKSVALSPRSY 855
RESULT 3
US-08-945-749-2
; Sequence 2, Application US/08945749
; Publication No. US2002013880A1
; GENERAL INFORMATION:
; APPLICANT: GARDNER, Richard C
; APPLICANT: MacDIARMID, Colin W
; APPLICANT: HAY, Robert J
; APPLICANT: Auckland Uniservices Limited
; APPLICANT: New Zealand Pastoral Agriculture Research Institut
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE
; FILE REFERENCE: 08/945,749
; CURRENT APPLICATION NUMBER: US/08/945,749
; EARLIER FILING DATE: 1998-01-12
; EARLIER FILING DATE: PCT/NZ96/0035
; EARLIER FILING DATE: 1996-05-01
; EARLIER FILING DATE: NZ 272039
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-08-945-749-2
Query Match 34.6%; Score 1666; DB 2; Length 858;
Best Local Similarity 40.9%; Pred. No. 2.8e-106;
Matches 401; Conservative 136; Mismatches 221; Indels 233; Gaps 35;
Qy 39 DSEDELELSESEFV-----KSEKQQQHHQBITSDNAKPLTRKSGSIKKSINLTDKD 93

9 DSSDDIPRASKYVDNTAASKCKYKPLENYRQ---YSDAQPIRE---AALKVDETKDS 62
 QY RITNPMSLSCDDTINSCHKVRY---NNS-SLRGDPYLVNDTNDSTNTHHTLAIP 148
 Db 63 R--HKFSSNGE---NSGVNGGVYKVTIWTSGRDP--EGEAEAVKR YOLRSAL- 114
 QY 149 IPIPTPIITANKSR---RKSOLENPPLIKKK---TIGRNNNSNFED 191
 Db 115 -----LSSNARPSRLAKSETHOKO IHVESIAPSLPKNAALRGHDTALPACTSSRCNL 168
 QY 192 LVSPMTKNTN-----DSEDITNT-----STTANHMLGIGATT 226
 Db 169 EASSARTPTTSARKASLVSAIFETSAESHGTHPKQAKLKRRTYSTISHS--SVNPTTL 226
 QY 227 GVGTVGTATATATAAG-----RRPSRSSIDSEADSHASRSSQETEDVCFPMVG- 276
 Db 227 -----LTRTASOKSEMGNDTRRIKPLRMDSRVSPHSEI--SQASRDSQETEDVCFPMFL 280
 QY 277 DHIRVNGIDPOEIDETIREREERAYLQKMIATAKILRIDEFONLKSNNNTTSCASRH--P 333
 Db 281 LHRVNGVDPOELEEA-----QTSARNLNSLANHORHSERT 318
 QY 334 YHHHSNNKKNNGDGGGGSSMAALKYTPKNILKTLRSRPTTHENSSSSEIYELKTKQO 393
 Db 319 YNHTQDQGFYNSASSTGSS--AALKVTPS-----ISR---TLEKNCVNEYVSENNES 368
 QY 394 PPKYDDQLSLTSTSTSGSGSQVKFGARISDGLNGG-----SLPD 437
 Db 369 --VREDRKPDL-----HPDVTFRGRIKGEKEGNDSSYRAYVTLONTYQIPS 415
 QY 438 RFLSFHSESETHIADIPSLVSPGQSVRDLPFRNGEBETWLDCTCPTDSEMKHLAKAFGI 497
 Db 416 RF5FPFSESDTHASDIPSLISEGOTFYELFKGODPTWLDSCPTDDENEC TAKTFGI 475
 QY 498 HPLTAEDRMQETREKVELPKSYFYFCPTHTEADKESDYLEPINVYIVVFHDGILTFHP 557
 Db 476 HPLTAEDRMQETREKVELPKSYFYFCPTHTEADKESYLEPINVYIVVRSGLVTFHP 535
 QY 558 SPISHPANVRVRQRUDYDVSADMVLYALIDEITDGPVPIHGIEYDAIDAIDAVFTA 617
 Db 536 DPISHCANVRVRQRUDYVSNSDMVLYALIDITDSPAPIOYIEYDAIDSIDSVFTM 595
 QY 618 RDTDPSEMLQRIQESRKRVTMLRLLSGKADYIKHFAKQOEANSSSGYQRO---YNL 674
 Db 596 RMDPAMMLQRIQESRKRVTMLRLLSGKADYIKHFAKRCQDETNGIGPVLKSOQTNMVL 655
 QY 675 QQQQQQAAAAPPPPIPIITSPINSTLANSLGTSOGGVGGVGNPFGNPTGNNTNTNTNT 734
 Db 656 QAQE-----NQWQ-----NNSN-----668
 QY 735 GSPSPQQQQQHQITNKSPFPDAPRADIALYGDIOHITMFOHLAVEKTFSSHS 794
 Db 669 -----NQISLH--STNGTSTQSPGDIALYGDIOHITMFOHLAVEKTFSSHS 718
 QY 795 NYLAOLOVSEFNNSNNKITEMPSKITLGTMLPLVNLVCLGKQVVRCEGCTNLGHPFC 854
 Db 719 NYLAOLOVSEFNNSNNKITEMPSKITLGTMLPLVNLVCLGKQVVRCEGCTNLGHPFC 777
 QY 855 IGVGLFIITGSPFPAQWLKLLNNSI-----EGQNGNRPi-----PHUSSERSI 900
 Db 778 ILGVLALLAVISWFLASWIKKIDDPATLNEAGSGAKSVISSPLPERKDFNDSXNGN 837
 QY 901 RSLGLKHKGNKSIITSPBNKY 921
 Db 838 ARVGVR-----KSTVSLPSRY 854

RESULT 4
 US-10-128-714-3563
 : Sequence 3563, Application US10128714
 : Publication No. US20030119013A1
 : GENERAL INFORMATION:
 : APPLICANT: Jiang, Bo
 : APPLICANT: Hu, Wenqi
 : APPLICANT: Tishkoff, Daniel
 : APPLICANT: Zamudio, Carlos
 : APPLICANT: Eroshkin, Alexey M
 : APPLICANT: Lemieux, Sebastian M

305 -----IVEKLQNEENETRFPGFSESQSTVHAAELGDLVLPD 342

QV 464 SVRLDFRNGEE--TWLDCCTPTDSEKMLAKAFGHPHTAEDIRQETREKVELFKSYT 521

Db 343 TRDLFQLGPEGVWMLDVLNPTTEBEVAALSRAFSTHPLATEDILTQEAEREKVELFKQY 402

QV 522 FVCFHTF-EADKESDYLEPINVYVPHDGIUTPHFSPISHPANVRVRQRQDVDS 580

Db 403 FVCFHTFVQLDKTDSEFNPVNFVHVRDGVLSFSTENPHAAVNRKRGKLRDVSLS 462

QV 581 ADMLCYALIDEITDGFAPVHIGIEYADAEADVFTARDTDFSSMLQRIGESRRKVTLM 640

Db 463 SDWICYAMIDIDVDSFGVIRETEIESSAIEDLVPIARVDDPESFLPRIGLKKVNSLM 522

QV 641 RLSCGKADVIKFAKCOEANSSSGYQRYQVNLQOOQQA PPPPNPIITSPINSLNL 700

Db 523 RLSCGKADVIRGSKRNE----- 541

QV 701 NSLGTSTGGGVGGVGNFPGNPTGNTNTNTNTTGTSPSPQOOQOQGRTKNSPPIPDARP 760

Db 542 -----QVSVT-----P 547

QV 761 RADIALYLGDIQDHIITMFONLLAYEKIFSRSHSNYLAQLQVESFNSNNKITEMFSKITL 820

Db 548 RGDIGLYGDIQDHRVTNMSNLAHPEKMLSRSHNYLAQLNVTNLVGNHVNKVLKVTL 607

QV 821 IGTMLVPLNLVTGLFGNNVRVPEGGNTLGMWFFGIVGVLIIFIIGSFIPAQMW 873

Db 608 IATMLVPMNLICGLFGNNVRVPEGQGLGMWFFGIVGVIAAVIVLSGIAARYY 660

RESULT 6

US-08-945-749-3

Sequence 3, Application US/08945749

Publication No. US2002013880A1

GENERAL INFORMATION:

APPLICANT: GARDNER, Richard C

APPLICANT: MACDIARMID, Colin W

APPLICANT: HAY, Robert J

APPLICANT: Auckland Uniservices Limited

APPLICANT: New Zealand Pastoral Agriculture Research Institut

TITLE OF INVENTION: ALUMINUM RESISTANCE GENE

FILE REFERENCE: 08/945.749

CURRENT APPLICATION NUMBER: US/08/945.749

EARLIER FILING DATE: 1998-01-12

EARLIER APPLICATION NUMBER: PCT/AZ96/0035

EARLIER FILING DATE: 1996-05-01

EARLIER APPLICATION NUMBER: NZ 272039

EARLIER FILING DATE: 1995-05-01

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

LENGTH: 969

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-08-945-749-3

Query Match 17.88; Score 856; DB 2; Length 969;

Best Local Similarity 24.18; Pred. No. 4e-50;

Matches 270; Conservative 143; Mismatches 248; Indels 460; Gaps 29;

QV 11 STTNQIPRSDEVLDDHRNQITNDCAISDESELEKSESEVVKSE----- 58

Db 44 SDSRRPTQLLDHNLQHNHGOITDPQI-DSWQMLHESDSTNDIKSEDPKLGAFIDHR 102

QV 59 -----KQOHHQBITSD-NAPLTRKSGSSIKKKNLTKDKRITNPM-LSGGDDTINSCH 112

Db 103 PMSQPREGQSVSTVQPIKMFSTPSYKPKACLPASQVRSVLSLSPSE--LESML 160

QV 113 KRNVNYSRLKDFYLDKN-TDQNSTNNHTHIAIPIPIPTPIITNA-----NKSRK 165

Db 161 KRR-----KSVHKSFPVDSNPTDRQSNANNVDVVDV-ALMNVHVNNASTGVNDNSKRR 215

QV 166 SQLENLPLIKKTKTGRNNSNFENDLVSPMTKMTNDSSEDTNTSTANRKLIGATT 225

Db 216 -----KKR--GSDSSN-KN---SKSTSDSNDDEBYSN----- 244

QV 226 LGVGTGTATATATAAGRRPSKSIDSEADSHASRSQETEDVCFPM--VGDHVRNG 283

QV 761 RADIALYLGDIQDHIITMFONLLAYEKIFSRSHSNYLAQLQVESFNSNNKITEMFSKITL 820

Db 548 RGDIGLYGDIQDHRVTNMSNLAHPEKMLSRSHNYLAQLNVTNLVGNHVNKVLKVTL 607

QV 821 IGTMLVPLNLVTGLFGNNVRVPEGGNTLGMWFFGIVGVLIIFIIGSFIPAQMW 873

Db 608 IATMLVPMNLICGLFGNNVRVPEGQGLGMWFFGIVGVIAAVIVLSGIAARYY 660

RESULT 5

US-10-128-714-8563

Sequence 8563, Application US/10128714

Publication No. US20030119013A1

GENERAL INFORMATION:

APPLICANT: Jiang, Bo

APPLICANT: Tishkoff, Daniel

APPLICANT: Zamudio, Carlos

APPLICANT: Eroshkin, Alexey M

APPLICANT: Lemieux, Sebastien M

TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus

and

TITLE OF INVENTION: Methods of Use

FILE REFERENCE: 10182-018-999

CURRENT APPLICATION NUMBER: US/10/128.714

PRIOR FILING DATE: 2002-04-23

PRIOR APPLICATION NUMBER: US 60/285.697

PRIOR FILING DATE: 2001-04-23

PRIOR APPLICATION NUMBER: US 60/287.066

PRIOR FILING DATE: 2001-04-27

PRIOR APPLICATION NUMBER: US 60/295.890

PRIOR FILING DATE: 2001-06-05

PRIOR APPLICATION NUMBER: US 60/303.899

PRIOR FILING DATE: 2001-07-09

PRIOR APPLICATION NUMBER: US 60/316.362

PRIOR FILING DATE: 2001-08-31

NUMBER OF SEQ ID NOS: 8603

SOFTWARE: PatentIn version 3.1

SEQ ID NO 8563

LENGTH: 663

TYPE: PRT

ORGANISM: Aspergillus fumigatus

US-10-128-714-8563

Query Match 22.88; Score 1099; DB 4; Length 663;

Best Local Similarity 32.78; Pred. No. 3.5e-67;

Matches 292; Conservative 98; Mismatches 241; Indels 262; Gaps 22;

QV 4 SESVYQNSTTNQIPRSDEVLDDHRNQITNDCAISDESELEKSESEVVKSEKQOQH 63

Db 7 SEKYVCLTRESTVPE---LDDHFFQL-----DSPPRIATADI-SLSRQNTAQHAY 34

QV 64 HQBITSDNAKPLTRKSGSSIKKKNLTKDKRITNPMSLGGDDTINSCHKNRNYSLSR 123

Db 55 HQBITPQ-----RPDLSTQDALREAGSLSRDFEQALD 87

QV 124 KDFYKLD-----NTDQNSTNNH-----THLAIPPIPIPTPIITNANKSRK--SQ 167

Db 88 DRSKQDINTLGRFSDVDPNGVNRHGRWTRTHQEL-----ANMSRESSPSA 134

QV 168 LENLPLIKKTKTGRNNSNFENDLVSPMTKMTNDSSEDTNTSTANRKLIGATTG 227

Db 135 RESSPP-----NSVEAPADPRRRERANTLE-----SHAAPDLAILQRTVSG 176

QV 228 VGTGT-TATATATAAAGRRESSISDSEADSHASRSQETEDVCFPMGDHVRNGIDR 286

Db 177 ---GTHPRPRPTSNASAIRPQDQIOLPN-----DESCVPTTEQGRIPVIDY 222

QV 287 DBIDFTREEREAYLQKQIAKNILRIDEFQNLKKNITTSAGSRH---YHHSNNKK 343

Db 223 EBLEEFA-----LSRQKPSR--R-----RQSLSSQSGRPRVFDLRPLGRKS 266

QV 344 NNGDGGGSSMAALKYTPKNILKTLRSFEFTHENSSSSEYELKTKQPPKYDDQLS 403

Db 267 DVEGSKASSA-----DRSSDMDADLKTADKYANVVDK 304

QV 404 LTSSTSTSGSGSQVFGAGHISDNGGSLDRFSLPHSESETHIAPDIPSLVSPQC 463

US-09-864-408A-2266

Query Match 7.2%; Score 345.5; DB 3; Length 154;
Best Local Similarity 44.64; Pred. No. 6.8e-16;
Matches 79; Conservative 20; Mismatches 39; Indels 39; Gaps 6;

QY 351 GSSAAKYTPKNIKLTLSRPEPHNSSSSEIYELKTKQPPKYDQLSLTSSISS 410
DB 1 GSS-AAKYTP-----ISR-----TLEKNSVNYKSYNNES--VREDPDL----- 42

QY 411 TSGSGGVKFGANISDGLNG-----SLPDRFSLPHSSEETHPD 454
DB 43 -----HPDVTFGNKIEGEGNDSSVRYAYTLQNTVEQTPSRFSFSESDETVHSD 97

QY 455 IPLSLVPGQSVDLPFGNEETWMLDCTCTDSEKMLAVKCIHPLTAEDIRMQETR 511
DB 98 IPLSLSEQPFYELKGDPTWMLDCCPTDDEARCIANTTGILHPLTAEDIRMQETR 154

RESULT 8
US-10-425-115-321932
; Sequence 321932, Application US/10425115
; Publication No. US2004021472A1
; GENERAL INFORMATION:
; APPLICANT: Le Rose, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 321932
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MNT4577_56667C.1.pep
US-10-425-115-321932

Query Match 6.2%; Score 298; DB 4; Length 94;
Best Local Similarity 61.4%; Pred. No. 6.7e-13;
Matches 54; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

QY 573 LRDYVDSADWLCYALIDETDGPAPVHIGIYEVADAIEDAVFTARDTDFSSMLQIGES 632
DB 1 LRDYVLSDDWICYANIDDVDSFGPIKMDVIESEIAEDHVFVAKVDFGTFLPRLGDL 60

QY 633 RKXWTLNRLLSKADVKIRKFAKQCE 660
DB 61 RKXWTLNRLLSKADVKIRKFAKQCE 88

RESULT 9
US-08-945-749-8
; Sequence 8, Application US/08945749
; Publication No. US20020138880A1
; GENERAL INFORMATION:
; APPLICANT: GARDNER, Richard C
; APPLICANT: GARDNER, Richard W
; APPLICANT: MCDONALD, Colin W
; APPLICANT: HAY, Robert J
; APPLICANT: Auckland Uniservices Limited
; APPLICANT: New Zealand Pastoral Agriculture Research Institute
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE
; FILE REFERENCE: 08/945,749
; CURRENT APPLICATION NUMBER: US/08/945,749
; CURRENT FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: PCT/NZ96/0035
; EARLIER FILING DATE: 1996-05-01
; EARLIER APPLICATION NUMBER: NZ 272039
; EARLIER FILING DATE: 1995-05-01
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Synechocystis sp.

Query Match 6.2%; Score 298; DB 2; Length 387;
Best Local Similarity 23.0%; Pred. No. 4.8e-12;
Matches 108; Conservative 70; Mismatches 143; Indels 148; Gaps 18;

```

: RESULT 10
: US-11-097-143-7905
: : Sequence 7905, Application US/11097143
: : Publication No. US20050208558A1
: : GENERAL INFORMATION:
: : APPLICANT: Venter, J. Craig
: : APPLICANT: et al.
: : TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
: : TITLE OF INVENTION: ARRAYS FOR DETECTING EXPRESSION OF 10,000 OR MORE
: : TITLE OF INVENTION: DROSOPHILA GENES.

```

```

? FILE REFERENCE: C:\000728
? CURRENT APPLICATION NUMBER: US/11/097,143
? CURRENT FILING DATE: 2005-04-04
? PRIOR APPLICATION NUMBER: 60/157,832
? PRIOR FILING DATE: 1999-10-05
? PRIOR APPLICATION NUMBER: 60/160,191
? PRIOR FILING DATE: 1999-10-19
? PRIOR APPLICATION NUMBER: 60/161,932
? PRIOR FILING DATE: 1999-10-28
? PRIOR APPLICATION NUMBER: 60/164,769
? PRIOR FILING DATE: 1999-11-12
? PRIOR APPLICATION NUMBER: 60/173,383
? PRIOR FILING DATE: 1999-12-28
? PRIOR APPLICATION NUMBER: 60/175,693
? PRIOR FILING DATE: 2000-01-12
? PRIOR APPLICATION NUMBER: 60/184,831
? PRIOR FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: 60/191,637
? PRIOR FILING DATE: 2000-03-23
? NUMBER OF SEQ ID NOS: 43008
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 7905
? LENGTH: 1198
? TYPE: PRT
? ORGANISM: DROSOPHILA
? US-11-097-143-7905

Query Match          4.9%; Score 234.5; DB 6; Length 1198;
Best Local Similarity 19.7%; Pred. No. 5.7e-07;
Matches 185; Conservative 114; Mismatches 291; Indels 349; Gaps 41;

```

QY 57 SEKOOHQHQTSDNAKLTRSGSS-----TKKS-----87

Dd 37 SHOQQOQQQLLPHHHKDKWLAAGSSPLKPFYSHILOQKDQATATGPAALAAAEVAAATT 96

QY 88 -----NLTKDRIITNPWSGGDDT---INSGHKNRYNM-----SLSRKDFVLAKN 131

Dd 97 SANADNFSSLOTDASQDGGISLSGLCDRFPVASPNPNSNMTLAGTATAATTTTTNN 156

QY 132 TDDNSTNNHTHLAIPPIPT-----PILTANKSRKSOLENLPPLIKKT 179

Dd 157 NNNNTNNNNNNVEAKTVRPNGSNVIESVTMPSFANILPTHRSENECIDPALQKN 216

QY 180 IGRUNSNFTENDLSP-----MYKKNTDESIDT---TSYTAN--RMGLGTG-----222

Dd 217 PQNPNGNN--SSIIVPVVEYHQLKPLEVNSSTSVTSNPLSSTTAQLDLDPVOVGKDDCHI 275

QY 223 -ATTLLGVGTG-----TATATATAAGR--RPSRSSTI-----251

Dd 276 STTTTTPGSGSASGSGSGSGSIARTIGTATPTTTTMSNTANTPTRASSLUHSIELA 335

QY 252 -----DSBADSHARSQSOTTEEDVCFPMVGDIHVNV-----GIDFD 287

Dd 336 ASSCA PRAASPNSNHTSSASTTPOOQQOQHMGSGNHSGSNLSDDBSMSEDEFGLEID 395

QY 288 RTDFETREEREAYLOKANILRIDFO-----NLSKNNTTSG-----ASHH 332

Dd 396 DNGEMDQQATVQFISSLNISPPSQMLEQQOQSSPALAAGGNSSNNAASGSNNNSASGN 455

QY 333 PYVHHNSNNKNN-----GCGGGSSMAALKTY-----360

Dd 456 NTSSSSNNNNNNNDANDHVLTKPEHEYNAYTLQLAGGGSGGNQOHHSNHSNHNHQ 515

QY 361 -----PKNLKLTLSRFETHENSSESSEIEYLKTKQOPPY 396

Dd 516 QOQQOQQOQQOQHQQOQBHYQOQQOQNINANA-NQF-----NSSSYSVIYNPDQSYIPT 569

QY 397 KYDDLSTLSTSTSGSGSGGVKFGARISDCINGGS-----LPDRFLSHSEEE 448

Dd 570 GYQD--TTSSHSSQSGGGGG--GGGNLNGSGSGSAGGYMLLP-----611

QY 449 TTHAPDIPSLVSQSVVDLDFNGEBETMWLDCTCPTOSEMKMLAKAFGI-----497

Dd 612 -----QAASSSGNNGN-----PNAGHMSGSGVNGSGGAGGAGGN 648

QY 498 -----HPLTAEDIRMOETREKV--LFKSYVPVCPTHFEADKESDYLEPINVYTVVFHD 550

Dd 649 SGPNGNPGCTSATHCGGEVIDFKHLFEELPCVC-----GDKVSQ-----YHY 691

QY 551 GLTFHPFSIPHANVRVRVRLRDYDVYSADMCLYALIDEITDGFAPVTH-----GI 603

Dd 692 GLLTCE-----CKGFFKRTVQKKVYTCV-ABRSCH-IDKQRKRCPCYCRFKCLEVM 744

QY 604 EYEADATEDAVPTARDTDFSSMLOR-----IGESRKVMTLARLLSKGADVIMKFAKRQC 658

Dd 745 KLEA-VRAIRMGRGN-KFGPMYRDRARKLRQWRQRLAQALQALNSNGSPDIK-----795

QY 659 BEANSNGSGYQRVNLQOQQO-----QAPPNNPITPSINSTLNLSLGSTG 708

Dd 796 -PTPIPSGYQQAYPNNNIKEIQIPQVSSLTQSPDSSPSPI-----AIALGOVNASTG 847

QY 709 -----GGVGUGGIFGCPNPTGNNTNTNTNTTYS 736

Dd 848 GVIAETPWNAICTGSGGGGLN-GPSSVGNGSGNSNGSNGN 885

RESULT 11
US-08-945-749-7
; Sequence 7, Application US/08945749
; Publication No. US2002013880A1
; GENERAL INFORMATION:
; APPLICANT: GARDNER, Richard C
; APPLICANT: MacDIARMID, Colin W
; APPLICANT: HAY, Robert J
; APPLICANT: Auckland Uniservices Limited
; TITLE OF INVENTION: New Zealand Pastoral Agriculture Research Institut
; FILE REFERENCE: 08/945,749
; CURRENT APPLICATION NUMBER: US/08/945,749

RESULT 11
US-08-945
; Sequenc
; Publica
; GENERA
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; TITLE
; FILE R
; CURREN

; CURRENT FILING DATE: 1998-01-12
 ; EARLIER APPLICATION NUMBER: PCT/NZ96/0035
 ; EARLIER FILING DATE: 1996-05-01
 ; EARLIER APPLICATION NUMBER: NZ 272039
 ; EARLIER FILING DATE: 1995-05-01
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 380
 ; TYPE: PRT
 ; ORGANISM: Synechocystis sp.
 US-08-945-749-7

Query Match 4.48; Score 214; DB 2; Length 380;
 Best Local Similarity 19.28; Pred. No. 3.1e-06;
 Matches 84; Conservative 72; Mismatches 141; Indels 140; Gaps 14;

Qy	478	LDCTCPTDSE-----MKMLAKAFGIHPLTAEDIRMOETREKVELFSYF	522
Db	59	LDCEAYDTSVSWINIDGLNHTWQGLGEVFKLHPVALEDIVNVPQKVVYENHLI	118
Qy	523	VCPTHFEADKESDYLEPINVIYVFDHGLTFHFSF-ISHPANVRRVRQLDYV-DVS	580
Db	119	FISRMVTLQSSQTFISEQISFILGKH-YLLTIQEEPKYDCLFSVRERIEIKKGAIRQN	177
Qy	581	ADMCLYALIDEITDGFAPVHIGIEYEADEAVFTARDTDFSSMLQIGESRRKVTLM	640
Db	178	ADYLFYALIDAIDGFFV---MEYGLSVQSLQSEIISCPTNKSIAKHQLOQDOLLIR	234
Qy	641	RLLSGKADVIKMKFAKRCQBEANSSSGYYQRYNQLQQQQQAPPPPNPIITSPINSTLNL	700
Db	235	RAIWPQDRAI-----	244
Qy	701	NSLGTSGCGVGGVGINFGPNFTGNNTNTNTTGTSPSPQQQQOHHGITNKSFPIDARP	760
Db	245	NSL-----LRDGSLLSD---	257
Qy	761	RADIALYLDIQDHIITWFOHLLAYEKIFSRSHSNYLAQOVESFNSNNKITEMSKITL	820
Db	258	--EVRVLRCDYDHTIQLDMETRYDLASNTDIYLSV-----SNRMEINKTLTV	308
Qy	821	IGTMLVPLNLVTLFGMNVVRVCGEGTN---LGMFFG-IVGVLIIFIIGSIFPAQW---	872
Db	309	ISSEIFLTFIAGIYGNFN-PKSPWNPMLWYGVYVWVNLTVGGMMLYFFWRKG	367
Qy	873	WLKUNLSIEGQNGNR	889
Db	368	WFRNLND-----VEKGNR	380

RESULT 12
 US-10-156-761-9574
 ; Sequence 9574, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 9574
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-9574

Query Match 4.48; Score 213.5; DB 4; Length 386;
 Best Local Similarity 19.68; Pred. No. 3.4e-06;

Matches 86; Conservative 67; Mismatches 150; Indels 135; Gaps 13;

Qy	459	VSGQSVRLDFRNGEE-----TWMLDCTCPTDSEMKMLAKAFGIHPLTAEDIRMOETREKV	514
Db	62	VSPASLADTFRELDRSPGMAWIGLARPTSESLLSLAEEFDLHPLAVEDAMEAHQRPKL	121
Qy	515	ELFKSYFVFCFHT--FEADKESDYLEPINVIYVFDHGLTFHFSFISHPANVRRVRQ	572
Db	122	ERYGETLVVYLRARAYLDAPEEVDGE---LHVFGDPFVTVVRHGAAPDLSAVRRNEE	178
Qy	573	LADYVDSADMLCYALIDEITDGFAPVHIGIEYEADEAVFTARDTDFSSMLQIGES	632
Db	179	TPELLKLPPEAVLYALDAVDGYVFWVAGVQNDIDEIETEVEFRGD---PAVSRIYEL	234
Qy	633	RRKVMTLRLLSGKADVIKMKFAKRCQBEANSSSGYYQRYNQLQQQQQAPPPPNPIITS	692
Db	235	SREMYEFQR-----	246
Qy	693	PINSTLNLSTGTSGCGVGGVGINFGPNFTGNNTNTNTTGTSPSPQQQQOHHGITNKS	752
Db	247	PLVGLMLH-----SLMAGFAKYGTD-----	265
Qy	753	FPIDAPRADIALYLDIQDHI-----ITMFOHLLA-----YEKIFSRSHSNYLAQQLQ	801
Db	266	-----BELQRYLRDVAHRTHTSERVDGFRQALADILTVNATLVTOQNAEKRALA	316
Qy	802	VESFNSNNKITEMSKITLIGTMLVPLNLVTLFGMNVVRVCGEGTNLGMFFG---IVGV	858
Db	317	EAGFEQNEEI---KKISSWAAILFAPTLVGTIYGMNF---EHPMELGWSFGYPPAIGL	368
Qy	859	LIFIIGSFIP--AQWML	874
Db	369	MLGVCSLYVIFKRRGML	386

RESULT 13
 US-10-156-761-10644
 ; Sequence 10644, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 10644
 ; LENGTH: 371
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-10644

Query Match 4.28; Score 201; DB 4; Length 371;
 Best Local Similarity 19.78; Pred. No. 2.3e-05;
 Matches 91; Conservative 64; Mismatches 160; Indels 148; Gaps 13;

Qy	436	PDRLSFLPSESE-ETIHADIPSLV-----	SPQSVRLDFRNGBETW	476
Db	13	PSRPSMRKGSAYDTTRDFSTESAIVTDCAIYRDGARVATDRPLTPEAMRQVRDGGFV-	71	
Qy	477	WLDCTCPTDSEMKMLAKAFGIHPLTAEDIRMOETREKVELFSYFVCF---HTEADK-	532	
Db	72	WIGLHEPTEAFAGIAGEFGLHPLAVEDAVQAHQPKLEYDLSLTVFTKTHVYEHQDL	131	
Qy	533	-ESDYLEPINVIYVFDHGLTFHFSFISHPANVRRVRQRLDYVDVSDMLCYALIDE	591	
Db	132	TANSEVTEGVNCFTRDFFITVRHGGQSURLRHRHQDDPELLAGSVAVLHATDH	191	
Qy	592	ITDGFAPVHIGIEYEADEAVF-----TARDTDFSSMLQIGESRRKVTMTLRLLSG	645	

QY	620	TDFFSLQRI	GESREKVT	LMRL	LSGKAD	VIKMFAKRC	QEBEANS	SSGGYQ	QYQVNI	QOQOQ	679
Db	223	-----	-----	-----	-----	-----	-----	-----	-----	-----	238
QY	680	QAPPPPP	PIITS	PINST	LNLS	LGST	STGGV	GVGIN	FGPN	PTGN	TNTNT
Db	239	-----	-----	-----	-----	-----	-----	-----	-----	-----	251
QY	740	POOQOQ	HITN	KSP	PI	DAR	PRAD	IAL	YLQD	IOD	HIIT
Db	252	VVOR	-----	-----	-----	-----	-----	-----	-----	-----	304
QY	797	LAOLQ	VES	PN	SK	ITEM	FSK	ITL	IG	TM	VL
Db	305	VAORQ	-----	-----	-----	-----	-----	-----	-----	-----	854
QY	855	--IVG	VL	PI	IIG	SF	IP	--AOW	872		
Db	351	LALLA	MLG	FTLL	LYW	IF	RSK	W	372		

Search completed: June 13, 2006, 14:23:59
Job time : 187 secs

OM protein - protein search, using sw model

Run on: June 13, 2006, 13:59:19 ; Search time 200 Seconds
(without alignments)
2107.768 Million cell updates/sec

Title: US-10-018-105A-4

Perfect score: 4820

Sequence: 1 MSDESRYQNTNQPIPR.....LGLKKHGKNKSIISFPNKYE 922

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*
10: Geneseq2006s.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4820	100.0	922	3 AAB36515	Aab36515 Candida a
2	4814	99.9	922	5 ABP73838	Abp73838 Candida a
3	1679	34.8	859	2 AAW07872	Aaw07872 Aluminium
4	1666	34.6	858	2 AAW07873	Aaw07873 Aluminium
5	1099	22.8	663	6 ABJ25905	Abj25905 Aspergill
6	1099	22.8	663	6 ABJ25905	Abj25905 Aspergill
7	865	17.9	969	7 ADK62466	Adk62466 Disease t
8	821	17.0	228	5 ABP05970	Abp05970 Human ORF
9	345.5	7.2	154	5 ABP32160	Abp32160 Human gly
10	319.5	6.6	107	5 ABP06753	Abp06753 Human ORF
11	251	5.2	326	9 AEM92441	Aem92441 M. xanthu
12	234.5	4.9	1198	4 ABB60371	Abb60371 Drosophil
13	210.5	4.4	354	9 AEB40726	Aeb40726 L. pneumo
14	210.5	4.4	357	9 AEB37409	Aeb37409 L. pneumo
15	186.5	3.9	1072	5 ABB54963	Abb54963 Lactococc
16	182.5	3.8	373	4 AAG89814	Aag89814 C glutami
17	179.5	3.7	373	9 AED46971	Aed46971 Membrane
18	179.5	3.7	1172	5 ABB93308	Abb93308 C. albica
19	179	3.7	1199	4 ABB58274	Abb58274 Drosophil
20	175.5	3.6	1246	4 ABB61547	Abb61547 Drosophil

21	173.5	3.6	391	8 ADN47899	Adn47899 Thermococ
22	173	3.6	803	9 AEB53775	Aeb53775 Drosophil
23	170	3.5	1436	3 AAB18199	Aab18199 Plasmodiu
24	169	3.5	1468	4 ABB65329	Abb65329 Drosophil
25	168.5	3.5	816	6 AAE30116	Aae30116 Fruit fly
26	167.5	3.5	1458	10 AEF13870	Aef13870 Mouse NFA
27	166.5	3.5	1046	10 AEF38379	Aef38379 Dictyoste
28	166.5	3.5	1046	10 AEF38399	Aef38399 Dictyoste
29	166.5	3.5	2261	6 ABJ18914	Abj18914 Pathogen
30	166.5	3.5	2271	6 ABM72734	Abm72734 Staphyloc
31	166.5	3.5	2283	6 ABP56876	Abp56876 Staphyloc
32	166	3.4	688	5 ABP74039	Abp74039 Candida a
33	165.5	3.4	1977	7 ADE15647	Ade15647 Human str
34	165	3.4	2112	4 ABB60403	Abb60403 Drosophil
35	165	3.4	4498	4 ABB58595	Abb58595 Drosophil
36	164.5	3.4	2271	6 ABU16000	Abu16000 Protein e
37	161	3.3	366	4 AAG81142	Aag81142 Mycobacte
38	161	3.3	471	7 ADB74419	Adb74419 Mycobacte
39	159.5	3.3	1169	6 ABU43255	Abu43255 Protein e
40	159.5	3.3	1177	5 ABP40261	Abp40261 Staphyloc
41	159.5	3.3	1177	8 ADS06718	Ads06718 Staphyloc
42	159.5	3.3	1253	3 AAB19772	Aab19772 Human den
43	159.5	3.3	1253	5 ABP51785	Abp51785 Human den
44	159.5	3.3	1253	9 ADY28712	Ady28712 Human pho
45	159.5	3.3	2150	5 AAO22566	Aao22566 Wooden le

ALIGNMENTS

RESULT 1	
AAB36515	
ID	AAB36515 standard; protein; 922 AA.
XX	
AC	AAB36515;
XX	
DT	06-MAR-2001 (first entry)
XX	
DE	Candida albicans CaALR1 protein sequence.
XX	
KW	Candida albicans; yeast pathogen; identification; fungal; antifungal;
KW	CaKRE5; CaALR1; CaCDC24; diagnosis; fungicide; fungal infection.
XX	
OS	Candida albicans.
XX	
FN	WO200068420-A2.
XX	
PD	16-NOV-2000.
XX	
PP	05-MAY-2000; 2000WO-CA000533.
XX	
PR	05-MAY-1999; 99US-0132878P.
XX	
PA	(MYCO-) MYCOTA BIOSCIENCES INC.
XX	
PI	Roemer T, Bussey H, Davison J;
XX	
DR	WPI; 2000-687652/67.
DR	N-PSDB; AAC87954.
XX	
FT	New DNA encoding essential proteins of Candida albicans, useful for
FT	diagnosing fungal infections and to screen for clinical or agricultural
PT	antifungal agents.
XX	
PS	Claim 14; Fig 2A; 76pp; English.
XX	
CC	The present sequence represents the Candida albicans CaALR1 protein. The
CC	present invention describes the fungus-specific genes CaKRE5, CaALR1 and
CC	CaCDC24 isolated in the yeast pathogen C. albicans. The genes have
CC	antifungal and fungicide activity. The genes in C. albicans and are

CC useful as drug targets. Fragments of them are useful as probes and
CC primers for diagnosis of fungal infections, also as antisense and
CC ribozyme agents. Proteins encoded by the genes are used to screen for
CC their specific inhibitors which are potential antifungal agents for
CC controlling a wide range of fungi pathogenic on animals or plants.
CC Fragments of the proteins are also used to raise specific antibodies.
CC Higher animals do not contain genes closely similar to the C. albicans
CC genes, so antifungal agents that target them are less likely to be toxic
CC to humans
XX
SQ Sequence 922 AA;

Query Match 100.0%; Score 4820; DB 3; Length 922;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 922; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDESYQQNTTNPQIPRSDVLDHNRQITNDCAISDSEDELEKSELESEVVKSEKQ 60
DB 1 MSDESYQQNTTNPQIPRSDVLDHNRQITNDCAISDSEDELEKSELESEVVKSEKQ 60
QY 61 QQHQEITSNAKPLTRKSGSSIKKSNLTKDRITNPMLSGGDDTINSGHKNRYNMS 120
DB 61 QQHQEITSNAKPLTRKSGSSIKKSNLTKDRITNPMLSGGDDTINSGHKNRYNMS 120
QY 121 SLRKDFYKONTDNNSTNNHTLAIPIPITPIITNANKSRKSOLENLPLIKKTTI 180
DB 121 SLRKDFYKONTDNNSTNNHTLAIPIPITPIITNANKSRKSOLENLPLIKKTTI 180
QY 181 GRNNSNFENDLVSPMTKMTNDSIEDITNTSTANHKMLGIGATTLGVTGTTATATA 240
DB 181 GRNNSNFENDLVSPMTKMTNDSIEDITNTSTANHKMLGIGATTLGVTGTTATATA 240
QY 241 AAGRPRSSISDSEADSHSRSSQETEDVCPVVGVDHVRVNGIDFDEIFREREAA 300
DB 241 AAGRPRSSISDSEADSHSRSSQETEDVCPVVGVDHVRVNGIDFDEIFREREAA 300
QY 301 YLQKQMTAKNLIIRIDFQNLKNNNTTSGASRPHYHHNSNNKNGGSGGSSMAALKYT 360
DB 301 YLQKQMTAKNLIIRIDFQNLKNNNTTSGASRPHYHHNSNNKNGGSGGSSMAALKYT 360
QY 361 PNKILKTLSEFTHNSSSSEIEYELKTKQPPYKDYDQLSTSTSTSGSGSQVK 420
DB 361 PNKILKTLSEFTHNSSSSEIEYELKTKQPPYKDYDQLSTSTSTSGSGSQVK 420
QY 421 FGARISDINGSGLPDRFSLPHSESETHAPDTPSLVSPQSVRLDFNGEETWMLDC 480
DB 421 FGARISDINGSGLPDRFSLPHSESETHAPDTPSLVSPQSVRLDFNGEETWMLDC 480
QY 481 TCPTDSEKMLAKAFGIHPLTAEDIRHQETREKVELPKSYVFCFTFEADKESDYLEP 540
DB 481 TCPTDSEKMLAKAFGIHPLTAEDIRHQETREKVELPKSYVFCFTFEADKESDYLEP 540
QY 541 INVYIVFHDGILTFHFSPISSHANVRRVRLQDYDVSADMLCYALIDEITDGFAPVI 600
DB 541 INVYIVFHDGILTFHFSPISSHANVRRVRLQDYDVSADMLCYALIDEITDGFAPVI 600
QY 601 HGIEYEADAEVPTARDTDFSSMLQRTGESRRKVTMLRLLSKADIVKFAKQOE 660
DB 601 HGIEYEADAEVPTARDTDFSSMLQRTGESRRKVTMLRLLSKADIVKFAKQOE 660
QY 661 ANSSSGYQQNTNQQQQQQAPPPPPPIITSPINSTLNLSLGTSTGGVGVGGINFGP 720
DB 661 ANSSSGYQQNTNQQQQQQAPPPPPPIITSPINSTLNLSLGTSTGGVGVGGINFGP 720
QY 721 NPTGNNTNTNTTSGSPQQQQQHGINTKSPFIPDARPRADIALYGDIDQHIITMQ 780
DB 721 NPTGNNTNTNTTSGSPQQQQQHGINTKSPFIPDARPRADIALYGDIDQHIITMQ 780
QY 781 NLLAYEKIFSRSHSNYLAQLQVESFNNSNNKITEMFSKITLIGTMLVPLNLVTLGFMNVR 840
DB 781 NLLAYEKIFSRSHSNYLAQLQVESFNNSNNKITEMFSKITLIGTMLVPLNLVTLGFMNVR 840

DB 781 NLLAYEKIFSRSHSNYLAQLQVESFNNSNNKITEMFSKITLIGTMLVPLNLVTLGFMNVR 840
QY 841 VPBEGGNTLQWFFGVGLVPIIIGSFIPAQWMLKLNNSIEGQNNRPIFNHSSRSI 900
DB 841 VPBEGGNTLQWFFGVGLVPIIIGSFIPAQWMLKLNNSIEGQNNRPIFNHSSRSI 900
QY 901 RSLGLKHKGNKSIISFPNKE 922
DB 901 RSLGLKHKGNKSIISFPNKE 922

RESULT 2
ABP73838
ID ABP73838 standard; protein; 922 AA.
XX
AC ABP73838;
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential protein SEQ ID NO 7675.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX
OS Candida albicans.
XX WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
DR WPI: 2002-566694/60.
DR N-PSDB; AB232388.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
PS Claim 44; SEQ ID NO 7675; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthesis, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for

CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
XX
50 Sequence 922 AA;

Query Match 99.9%; Score 4814; DB 5; Length 922;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 921; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSDSYQNSTNQIPRSDVLDHRNQITNDCAISDESELELSESEVVKSEKQ 60
Db 1 MSDSYQNSTNQIPRSDVLDHRNQITNDCAISDESELELSESEVVKSEKQ 60
Qy 61 QQHQETSDNAKPLTRKSGSIKKSNLTKDKRITNPKSLSGGDDTINSGHKNRYNMS 120
Db 61 QQHQETSDNAKPLTRKSGSIKKSNLTKDKRITNPKSLSGGDDTINSGHKNRYNMS 120
Qy 121 SLRDKFYLKNTDNDSTNNHTLAIPITPIITNANKSRKSKOLENPLPKKTTI 180
Db 121 SLRDKFYLKNTDNDSTNNHTLAIPITPIITNANKSRKSKOLENPLPKKTTI 180
Qy 181 GRNSNFENDLVSPMTKKTNWSEDTNTSTNHNKLGICATTLGVTGTTATATA 240
Db 181 GRNSNFENDLVSPMTKKTNWSEDTNTSTNHNKLGICATTLGVTGTTATATA 240
Qy 241 AAGRRPSRSDSEADSHASRSQTEEDVCFPMVGDHVRVNGIDFDEIFREEREA 300
Db 241 AAGRRPSRSDSEADSHASRSQTEEDVCFPMVGDHVRVNGIDFDEIFREEREA 300
Qy 301 YLQKMTAKNLRIDFQNLKNNNTSGASRPHYHHSNNKNGGSGGSAALKYTI 360
Db 301 YLQKMTAKNLRIDFQNLKNNNTSGASRPHYHHSNNKNGGSGGSAALKYTI 360
Qy 361 PKNLLKTLRPFTHENSSESEIYELTKQPPYDDQLSLTSTSTSGSGSQVK 420
Db 361 PKNLLKTLRPFTHENSSESEIYELTKQPPYDDQLSLTSTSTSGSGSQVK 420
Qy 421 FGARI SDGNGSLPDRPSLPHSESETHAPDIPSLVSPQSVRLDFNGEETWLD 480
Db 421 FGARI SDGNGSLPDRPSLPHSESETHAPDIPSLVSPQSVRLDFNGEETWLD 480
Qy 481 TCPTDSEKHLAKAFGHPHTAEDIRMQETREKVELFKSYVFCFTHFEADKESDYLEP 540
Db 481 TCPTDSEKHLAKAFGHPHTAEDIRMQETREKVELFKSYVFCFTHFEADKESDYLEP 540
Qy 541 INVYIVPHDGLTFHPSLISHPANVRVRQLRDYVDSADMLCYALIDEITDGPAPVI 600
Db 541 INVYIVPHDGLTFHPSLISHPANVRVRQLRDYVDSADMLCYALIDEITDGPAPVI 600
Qy 601 HGIEYADAEIVFTARDTDFSSNLQIGESRRKVTMLRLSLGKADVIKFAKRCQE 660
Db 601 HGIEYADAEIVFTARDTDFSSNLQIGESRRKVTMLRLSLGKADVIKFAKRCQE 660
Qy 661 ANSSGGYQRYNQV 720
Db 661 ANSSGGYQRYNQV 720
Qy 721 NPTGNNTNTNTTTCSPSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 780
Db 721 NPTGNNTNTNTTTCSPSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 780
Qy 781 NLLAYEKIFSRSHSNYLAQLQVESFNSNNKITEMFSKILGTMVLPLVLTGFGNVR 840
Db 781 NLLAYEKIFSRSHSNYLAQLQVESFNSNNKITEMFSKILGTMVLPLVLTGFGNVR 840
Qy 841 VPGEGGNTLGMFFGIVGLFIIIGSFIFAQWLKLNNSIEQNGNNGNPIFNHSSRSI 900
Db 841 VPGEGGNTLGMFFGIVGLFIIIGSFIFAQWLKLNNSIEQNGNNGNPIFNHSSRSI 900

Db 841 VPGEGGNTLGMFFGIVGLFIIIGSFIFAQWLKLNNSIEQNGNNGNPIFNHSSRSI 900
Qy 901 RSLGLKKGKGNKSIISPPNKYE 922
Db 901 RSLGLKKGKGNKSIISPPNKYE 922

RESULT 3
AAW07872
ID AAW07872 standard; protein: 859 AA.

XX
AC AAW07872;

XX 23-FEB-1997 (first entry)

XX Aluminium resistance gene ALR1.

XX Aluminium resistance; ALR1; ALR2; ARH1; cation; tolerance;
XX magnesium transport.

XX Saccharomyces cerevisiae.

XX W09634959-A1.

XX 07-NOV-1996.

XX 01-MAY-1996; 96WO-NZ000035.

XX 01-MAY-1995; 95NZ-00272039.

XX (AUCK-) AUCKLAND UNISERVICES LTD.

XX Gardner RC, Macdiarmid CW, Hay RJM;

XX WPI; 1996-506161/50.

XX ALR1-homologues, aluminium resistance genes ALR1 and ALR2 - used in the
treatment of cation deficiency, cation toxicity, esp. heart disease.

XX Claim 6; Fig 5; 62pp; English.

XX ALR1 and ALR2 are isolated by selecting vectors which confer high
aluminium tolerance, or that complement knockout mutations in ALR1, ALR2
and/or ARH1. The genes have homology with bacterial genes responsible for
divalent ion uptake. The ALR1 and ALR2 are magnesium transport genes. The
isolated transport genes can be used in the treatment of any plant.
XX animal or microorganism disease which results from a cation deficiency,
by producing an accumulation of cations in a plant deficient to those
cations, or in plants consumed by animals deficient in those cations. The
genes may also be used in the treatment of diseases resulting from cation
toxicity, esp. manganese toxicity and esp. for the treatment of heart
disease. Note: The ALR1 and ALR2 gene sequences are claimed and stated as
given in Figure 6 and Figure 7 of the specification, however these
figures are missing from the specification

XX Sequence 859 AA;

Query Match 34.8%; Score 1679; DB 2; Length 859;
Best Local Similarity 41.5%; Pred. No. 6.2e-119;
Matches 405; Conservative 137; Mismatches 258; Indels 176; Gaps 31;

Qy 1 MSDSYQNSTNQIPRSDVLDHRNQITNDCAISDESELEL---KSELESEVVK 57
Db 1 MSSSSS---SSESSPYLSRS-----NSLANTWVSMKTEDHTGLYDHRQHPDSLVRH 49

Qy 58 EQQOQHQETSDNAKPLTRKSGSIKKSS-----NLTKDRITNPKSLSGGDDTINSGH 112
Db 50 QPPTLKKEKSTKSTPIKQKSGATYNSHYDVGVSFSGRMDPDEQGGMDTV-AHH 108

Qy 113 KNR-----NTNKKSSL-----RKDFYKNDTDDNSTNNHT---HLAIPITPIPTPII 156

Aluminium resistance; ALR1; ALR2; ARH1; cation; tolerance; magnesium transport.

Saccharomyces cerevisiae.

WO9634959-AL.

07-NOV-1996.

01-MAY-1996; 96WO-NZ000035.

01-MAY-1995; 95NZ-00272039.

(AUCK-) AUCKLAND UNISERVICES LTD.

Gardner RC, Macdiarmid CW, Hay RJM;

WPI; 1996-506161/50.

ARH1-homologues, aluminium resistance genes ALR1 and ALR2 - used in the treatment of cation deficiency, cation toxicity, esp. heart disease.

Claim 6; Fig 5; 62pp; English.

ALR1 and ALR2 are isolated by selecting vectors which confer high aluminium tolerance, or that complement knockout mutations in ALR1, ALR2 and/or ARH1. The genes have homology with bacterial genes responsible for divalent ion uptake. The ALR1 and ALR2 are magnesium transport genes. The isolated transport genes can be used in the treatment of any plant. animal or microorganism disease which results from a cation deficiency, CC by producing an accumulation of cations in a plant deficient to those CC cations, or in plants consumed by animals deficient in those cations. The CC genes may also be used in the treatment of diseases resulting from cation CC toxicity, esp. manganese toxicity and esp. for the treatment of heart CC disease. Note: According to the disclosure the ARL2 amino acid sequence CC comprises 860 amino acids, however the sequence given in Figure 5 CC comprises 858 amino acids. The ARL1 and ARL2 gene sequences are claimed CC and stated as given in Figure 6 and Figure 7 of the specification, CC however these figures are missing from the specification

XX Sequence 858 AA;

Query Match 34.68; Score 1666; DB 2; Length 858;

Best Local Similarity 40.98; Pred. No. 6.1e-118; Indels 233; Gaps 35;

Matches 401; Conservative 126; Mismatches 221;

Qy 39 DSEDELKSELESEV-----KSEKQHQHQBETSDNAKPLTRKSGSSIKKSNLTKD 93

Db 9 DSSDLPRSKVDNTAAASKTKYKLENYRQ---YSDAQPIRHE---ALALKVDETKDS 62

Qy 94 RTNPMELSGDDDTINSCHKRNY-----NMS-SLRKDFYKONTDSTNNHTLPI 148

Db 63 R--HKFSSNGE---NSGVNGGVYKTNISTSGRMDF--EGEAEEAVKRYQCSFAL- 114

Qy 149 IDPIPTITNANKS-----RKSQLENLPLIKK-----TIGRNNNNFEND 191

Db 115 -----LSSNARPSLAKSETHQKIHVESIAPSLPKNAALRGHDHALPAGTSSNRCNL 168

Qy 192 LVSPMTKRTN-----DSEDTNT-----STTANRMLGIGATTLL 226

Db 169 EASSARTFTTSARKASLVSAIFETSAESEHGTHPKQAKLRRTYSTISTHS--SVNPTLL 226

Qy 227 GVGCTGTATATAAAG-----RRPSRSSIDSEADSHASRSQETEDVCFPMVG- 276

Db 227 -----LRTTASQKSDMGNDYRIKPLRMDSRVSPHSEI--SQASRDSQETEDVCFPMFRL 280

Qy 277 DHRVNGIDFDEIDFETREEREAYLQKQIAKNILRIDFQNLKNTTSGASH--P 333

Db 281 LHRVNGVDPELEEYA-----QISNAERNLSLANHQHSERT 318

109 QURASAILTSNARPSRLAHSMHQRLQYVESNI-----HTPKDVGVKRDYTWSSSTA 161

Qy 157 THANKSRKSOLENPLPLIKKTYGRNNSNFENDVLSPMTKKTNDSEDTNT----- 210

Db 162 SSGNKS-KLSASSASPITKVR-----KSLVSPVLPIHESKSDTHSKLAKPKK 210

Qy 211 ---STTANRMLGIGATTLCGVGTATATA---TATAAAGRPRSPRSSIDSEADSHASRSQ 264

Db 211 RYVTSYSHSSIN-PAVLLTKTSQSDADDTLERKPVMTASPDSDV--SQASRDSQ 268

Qy 265 ETEEDVCFPMVG--HIRVNGIDFDEIDFETREEREAYLQKQIAK-NILRIDFQNLK 322

Db 269 ETEEDVCFPMVQLHTRVNGIDFDELEEYAFANAE---KSQFLASLVQENQKYSVSG 325

Qy 323 -----NNTTSGASRHPYHHHNNNNKNGGSGSSMAALYTPKNIKLLKTLRSFETH 376

Db 326 DIGTSTSTSGSS-----AALAYTPR--VSQGTGKSESTN 359

Qy 377 ENSSSSEIYELKTKQPPYKYDQLSLTSTSTSGSGGQVFPFG-ARISDINGISGL 435

Db 360 ET-----EIHEKKEHEKIKPSLHPGISFGKNKVEGENENIPSPNDPAVCSYQDTDFQI 414

Qy 436 PORFSLPHSESETHAPDIPSLVSPQSVRLFRNGEETWMLDCTPTDSEKMLAKAF 495

Db 415 PWRFPSCSESETHVADIPSLVSEQTFYELFRGSEPTWMLDCSPTDDEKACIAKAF 474

Qy 496 GHPLTADIRMQETREKVELPKSYVFCFHTFADRESESDYLEPINVYLVFHDGLTP 555

Db 475 GHPLTADIRMQETREKVELPKSYVFCFHTFADRESESDYLEPINVYLVFHDGLTP 534

Qy 556 HFSPLSHPANRRVRQLRDYDVNSADMLCYALIDETDQFAPVIRHIGYEADATEDAVF 615

Db 535 HFGPLSHCANRRVRQLRDYVNVNSDMLCYALIDDTSDTSPAVTQISYEADATEDAVF 594

Qy 616 TARDPFSMLQIGESRRKVTMLRLSGKADVIKFAKQCEANSSGGYQVQNLQ 675

Db 595 MARDMDFAHLQIGESRRKVTMLRLSGKADVIKFAKQCEANGI----- 643

Qy 676 QQQQQAPPPPIITSPINSTLNLSLGTSTGGGVGGVGNFGNPNTGNTNTNTVTG 735

Db 644 -----GPALTQIN-TANLQARQNA-----SHIKNNSTTVPNY 678

Qy 736 SPSPPPQQQOQHGINTKSPFPDARPRADIALYLDIODHIITWQNLAYEKIFSRSHN 795

Db 679 APTTSQ-----PRGDIALYLDIODHLLTWFQNLAYEKIFSRSHN 720

Qy 796 YLAQLQVESFNSNNKITEMFSKITLIGTMLVPLNLTGLFQANVRVPGEGTNLGFEGI 855

Db 721 YLAQLQVESFNSNNKVTETMLGKVTHTIGTMLVPLNLTGLFQANVRVPG-NSSIAMWFI 779

Qy 856 VGVLIIFIIGSFPAQWLKLLNNSI---EQQNNRPIPN-----HSSRRSIRSLGLKK 907

Db 780 LGVLLALLAVLWFLASVYIKRIDPPATLINEAESGAKSVISSFLPKRKNRKNRKNINV 839

Qy 908 HGG--NKSIIISPNKY 921

Db 840 RAGPSNKSVASUPSKY 855

RESULT 4

AAW07873

XX AAW07873 standard; protein; 858 AA.

XX AAW07873;

XX AAW07873;

XX 23-FEB-1997 (first entry)

XX Aluminium resistance gene ALR2.

PR	31-AUG-2001; 2001US-0116362P.	
XX		
PA	(ELIT-) ELITRA PHARM INC.	
XX		
PI	Jiang B, Tishkoff D, Zamudio C, Ershkin AM, Hu W, Lemieux SM;	
XX		
DR	WPI; 2003-0931124/08.	
XX		
PT	Disclosure; Page: 175pp; English.	
PT		
PT	New purified or isolated nucleic acids of essential genes of Aspergillus	
PT	fumigatus, useful for treating or preventing infections by A. fumigatus,	
PT	or for treating a non-infectious disease in a subject e.g. cancer.	
XX		
PS		
XX		
CC	The invention relates to novel purified or isolated nucleic acids of	
CC	essential genes of Aspergillus fumigatus. The isolated nucleic acids of	
CC	the invention are used to treat or prevent infections by a pathogenic	
CC	organism such as A. fumigatus, to treat a non-infectious disease in a	
CC	subject (e.g. cancer), to prevent or contain contamination of an object	
CC	by A. fumigatus, or to prevent or inhibit formation on a surface of a	
CC	biofilm comprising A. fumigatus. The polynucleotides are useful for	
CC	expressing recombinant protein for characterisation, screening or	
CC	therapeutic use, as markers for host tissues in which the pathogenic	
CC	organisms invade or reside, for comparing with the DNA sequence of A.	
CC	fumigatus to identify duplicated genes or paralogues having the same or	
CC	similar biochemical activity and/or function, for comparing with DNA	
CC	sequences of other related or distant pathogenic organisms to identify	
CC	potential orthologous essential or virulence genes, for selecting and	
CC	making oligomers for attachment to a nucleic acid array for examination	
CC	of expression patterns, for raising anti-protein antibodies, as an	
CC	antigen to raise anti-DNA antibodies or to elicit another immune	
CC	response, and for identifying polynucleotides encoding the other protein	
CC	with which binding occurs or to identify inhibitors of the binding	
CC	interaction. The polypeptides may be used to raise antibodies or to	
CC	elicit immune response, as a reagent in assays designed to quantitatively	
CC	determine levels of the protein in biological fluids, as a marker for	
CC	host tissues in which pathogenic organism invade or reside, and to	
CC	isolate correlative receptors or ligands in the case of virulence	
CC	factors. This sequence represents a protein of one of the essential genes	
CC	of Aspergillus fumigatus of the invention	
XX		
SQ	Sequence 663 AA;	
	Query Match 22.8%; Score 1099; DB 6; Length 663;	
	Best Local Similarity 32.7%; Pred. No. 9.7e-75;	
	Matches 292; Conservative 98; Mismatches 241; Indels 262; Gaps 22;	
QY	4 SESYQNSTNQIPRSDEVLDHRRNQTDCAISDESELELKSBLSESVKSEVQQOH 63	
DB	7 SEKMKLRFSTVPB---LDHRRQL-----DSFPREATADI-SLRQNTAQHY 54	
QY	64 HQBITSDNAKPLTRKSGSSIKKKSNTDKDRINPKMSLGGDDTINSGHKRNYYNASSLR 123	
DB	55 HQETPQ-----RPDLLSIQDLREAGSLSRDPEQAILD 87	
QY	124 KDPLYKD-----NTDDNSTNH-----THLAIPIPIPTPIITNANKSRK--SQ 167	
DB	88 DDRSGKDINTLGRFSVPDGNVRRGRTWSKTHQEL-----ANNSRESSEA 134	
QY	168 LEMLEPLIKKTKTGRNINNPENDLVSPMTKMTNDSEDITNTSTTTANMKLGATG 227	
DB	135 RSSSPPE-----NSEVAFADPRRREANTLE-----SHAAPDLBAILQRTVSG 176	
QY	228 VGTGT-TATATATAAAGRRPSRSDISEADSHASRQSETEDEVCFPMVGDRVRNGVD 286	
DB	177 ---GTHPRPPTFSNAAIRPOPDQIQLEFN-----DESCVPTTEQGRPIVDY 222	
QY	287 DEIDFIREEREAYLQKQMTAKNILRIDEFQNLKNNTTSGASRHP---YHHHSNNKK 343	
DB	223 ELEEFEVA-----LSRQMKPSTSR-----KQSLSSQSGRPRVFDLRPLGRKS 266	

New purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*, useful for treating or preventing infections by *A. fumigatus*, or for treating a non-infectious disease in a subject e.g. cancer.

Disclosure; Page: 175pp; English.

The invention relates to novel purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as *A. fumigatus*, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or inhibit contamination of an object by *A. fumigatus*, or to prevent or inhibit formation on a surface of a biofilm comprising *A. fumigatus*. The polynucleotides are useful for expressing recombinant protein for characterisation, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of *A. fumigatus* to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organism invade or reside, and to isolate correlative receptors or ligands in the case of virulence factors. This sequence represents a protein of one of the essential genes of *Aspergillus fumigatus* of the invention

Sequence 663 AA;

Query Match 22.8%; Score 1099; DB 6; Length 663;
Best Local Similarity 32.7%; Pred. No. 9.7e-75;
Matches 292; Conservative 98; Mismatches 241; Indels 262; Gaps 22;

QY 4 SESYQNSTNTPRSDVLDHNRINOTNDCAISDEDELKSELESEVKVSEKKQQOH 63
||| : ||| : ||| :
DB 7 SEKYNCILTRFSTPVE---LDRHQFL-----DSRPRIEATADI-SLSQNTAOAHY 54
||| : ||| : ||| :
QY 64 HOETTSNAKPITRKSGSIKKKNLTDXKITNPMSLGDDTINSGHKNNRYNNSSLR 123
||| : ||| : ||| :
DB 55 HQETPO-----RPDLISIQDALREAGSLSDFEQAILD 87
||| : ||| : ||| :
QY 124 KDFYLKD-----NTDNNSTNNH-----THLAIPPIPIPTIITANKSRKK--SQ 167
||| : ||| : ||| :
DB 88 DRSCKDKINTLGRFRFSVDPNGNVRHGRTWSRTHOEL-----ANNRESPPSA 134
||| : ||| : ||| :
QY 168 LENLPPLIKKKTIGRNNSNNPDNLVSPMTKMKTNDSEIDTWTSTYANHKMLGICATTGL 227
||| : ||| : ||| :
DB 135 RSSFP-----NSVEAPADPRERANTLE-----SHAAPDALEQLRTVSG 176
||| : ||| : ||| :
QY 228 VTGT-TATATATAAAGRPRSISIDSEADSHASRSQTEDVCFFPMGDHRNVCIOP 286
||| : ||| : ||| :
DB 177 ---GTHPRAPTFSNAIRPQCIOLEPN-----DESCPTYEQGPAPIVDY 222
||| : ||| : ||| :
QY 287 DEIDEFREEREAYLOKQIAKNIKLIDEPONLSKNNTTSCASRHPP-----YHHHSNNKK 343
||| : ||| : ||| :
DB 223 EBLEEFVA-----LSRQMKPSTRS-----KQSLLSQSRGPRFYDLRLPGLRS 266
||| : ||| : ||| :
QY 344 NNGDGGGSSMAALKTYPRNKLYTLRSRPFTHENSSESSEELYELTKQPYPKYDDQLS 403
||| : ||| : ||| :
DB 267 DYEGKHSSSA-----DKSSOLDMDALKTAKYVANVVDKMD 304
||| : ||| : ||| :
QY 404 LTSSTSTSGSGSQVFKGARISDINGGSQPORFLPHSSEBETHADPIDSLSPSQ 463
||| : ||| : ||| :

Db 305 -----IWEKIQNENETRFGEFFSSSQSTVHAAELGDLVLPD 342
Qy 464 SVRLDFRNGEE--TWLDCCTPTDSEKMLAKAFGIHPLTAEDIRMOETREKVELFKSY 521
Db 343 TFDLPLQGLGEGVWMLDLNPTFEVVALSRAFSIHPLTIEDILITQAREKVELFKQY 402
Qy 522 FVCFHTP-EADKESDYLEPINVIYVFDHGLITVHFSPISHPANVRVRLQRYDVVS 580
Db 403 FVCFHTFVQLDKTDRFEPNPNVYVFRDGLVSEFTENPHAAANVRKIKGLRDYVSL 462
Qy 581 ADWLCVALIDEITGDFAVTHGIEYEDADEVFTARDVDFSSMLQIGESRRKRWTL 640
Db 463 SDWICYAMIDIVDFGVVREIESEHEDUFIARVDVFESFLPRIGGLRKRVSLM 522
Qy 641 RLSSKADVIMKFAKROBEANSSGVYQRYNLOQQOQQOQPPPPPIITSPINSTL 700
Db 523 RLGGKADVIRGSKRNE----- 541
Qy 701 NSLGTSTGGGVGGVGNFGNPTGNNTNTNTTNGSPSPQOQOQHGHTKSPFPIDARP 760
Db 542 -----QYSVT-----P 547
Qy 761 RADIALYLGDIQDIHITMQLLAYEKIFSRSHSYLAQLQVESFNNSKRNITEMSKITL 820
Db 548 RGDIGLYGLDIQDIHVTVMANLAHPEKMLSRSHNTYLAQLNVTNLVGNHVKLSKVTL 607
Qy 821 IGTMLVPLNLTGLFGNNVRVPEGGTNLGWFPGVGVGLIFIIIGSEPIFAQW 873
Db 608 IATMLVPMNLICGLFGNNVRVPEGQGLGWFPGVGVIAAVLVLSGAARY 660

RESULT 7

ADK62466
ID AUK62466 standard; protein; 969 AA.
XX
AC ADK62466;
XX
DT 06-MAY-2004 (first entry)
XX
DE Disease treating protein complex-derived protein #359.
XX
KW protein complex; drug target; diagnosis.
XX
OS Unidentified.
XX
FN EP1338608-A2.
XX
PD 27-AUG-2003.
XX
PF 20-DEC-2002; 2002EP-00102902.
XX
PR 20-DEC-2001; 2001EP-00130253.
XX
PA (CELL-) CELLZOME AG.
XX
PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
PI Marzoch M, Grandi P, Krause R, Kruse U, Nerino A, Bauch A;
PI Michon A, Leutwein C, Rick J;
XX
DR WPI; 2003-638460/61.
DR N-PSDB; ADK62467.
XX
PT New proteins and protein complexes from eukaryotes, useful as targets in
PT drug screening, or in diagnosing or screening for the presence of a
PT disease or disorder, or a predisposition for developing a disease or
PT disorder in a subject.
XX
PS Disclosure; SEQ ID NO 717; 13pp; English.
XX
CC The invention relates to novel protein complexes comprising a first and a

second protein, or its derivative, fragment, homologue or variant. The
proteins are selected from given protein complexes, which are not defined
in the specification. The variants are encoded by nucleic acids that
hybridize to the nucleic acids encoding the proteins under low stringency
conditions. The protein complexes are useful as targets for an active
agent of a pharmaceutical. These protein complexes are particularly
useful as drugs targets for the treatment or preventing of a disease or
disorder. The complexes and methods above are useful in diagnosing or
screening for the presence of a disease or disorder or a predisposition
for developing a disease or disorder in a subject. These are also useful
in screening for a drug for treatment or prevention of a disease or
disorder. The molecule that modulates the amount, activity or protein
components of the complex is useful for the manufacture of a medicament
for the treatment or prevention of a disease or disorder. This sequence
corresponds to a protein of the invention. (Note: the sequence data for
this patent did not form part of the printed specification but was
obtained from the EPO in electronic format).

XX Sequence 969 AA;

Query Match 17.9%; Score 865; DB 7; Length 969;
Best Local Similarity 24.2%; Pred. No. 1.3e-56;
Matches 271; Conservative 143; Mismatches 247; Indels 460; Gaps 29;

Qy 11 STTNOPIPRSDVLDHRRNQTNDCAISDSEDELELSESESEVVKSE----- 58
Db 44 SDSRRPTQLLDHNLQHNHQTIDTQOI-DSWGLHESDSTNDIIXKSDPSLKGAFIDHR 102
Qy 59 ----KQOQHQTSD-NAKPLTRKSGSSIKKSNLTKDKRITNPMG-LSGDDTTINGH 112
Db 103 PMSQREGFGSVSTVQPPQPKFSTPSYKKPAGLRPSQNRSLVSLSPSE--LESWL 160
Qy 113 KRNNTNMSSLRKDFYLDN--TDDNSTNNHTLAIPIPPIPTITNA-----NKRKR 165
Db 161 KXR----KSVKSFVDENSTDRQSNANDVDVYD-ALMNVNNAATCYVNDKRRK 215
Qy 166 SQLENPLIKKTKTGRNNPNFENDLVPMYTKMTNDSEDITNTSTANMKLGIGATT 225
Db 216 -----KKR--GSDSSN-KN--SKSTSSDNDDEYNS----- 244
Qy 226 LGVGTGTATATATAAGRPSPSSIDSEADSHASRSOETEDVCPM--VGDHVRNG 283
Db 245 -----RPSS-----LSSNSSLDDVCLVLDDEGSEVPKAW 275
Qy 284 IDPDEIDPIREREAYLQKQIAKNIILIDEFONLSKNNTSGASHPHYHHNNKK 343
Db 276 PDCTVLEEPSKEETER--LSQAI-----QDAEAPHQYDEDEE 312
Qy 344 NNGDGGGSSMAALAYTPKNIKLTLSRPEFTHESSSEIVEIKTKQPPKYVDQLS 403
Db 313 D-----GTSNEDGILFSKPIVNIIDVFNLRVNTENKNGRLRPRIAPWHLIQPM 367
Qy 404 LTSSTSTSGSGSGGVKFGGARISDINGGSL-----PD--RPSLFSES 446
Db 368 VL-----GSNSTKDSKRSIQSLQDNLVGNRIQYPPHIISNNPHEFTYFRVDL 418
Qy 447 EETTHAPDIPSLVSPQSVRDLP-----ADKES----- 469
Db 419 DSTVHSPTISGLLQPGQFDLFAVSYSDNSAGHKTKTPNSPTPGIKAEVTSOLQLT 478
Qy 470 -----RNGEET--WWLDCCTPTDSEKMLAKAFGIHPLTAEDIRMOETREKVE 515
Db 479 AKNPETLSSNSVANTIEDVPPFMDVSNPTEEMKILSKAFGIHPLTTEDIFLGEVREKE 538
Qy 516 LPKSYTVVCPHPTF-----ADKES----- 534
Db 539 LFRDYTLICFRSDIVAEKRVRRRREKQESATLDHESISRRKRSQAYGATMNSNANN 598
Qy 535 ----- 534

Db 599 NSTSNASRWLPSILARRRSSANRTTNTSSSYKRVKSEKKKHEENKFKKSGDRH 658
QY 535 ---EDYLEPINVIYVPHDGLTFHFSPISHPANVRRVRQLRDYVDVSDMLCYALIDE 591
Db 659 KPEGELEPLNVIYVFTGVTFHFAPTHPINVRRARLKDYLNTVTSWIAVALIDD 718
QY 592 ITGPAFVINGIEYADAEIVAFVTFARDTFSS----- 624
Db 719 ITDAFAPMBLIEDEVEIEDAILKMHQSDSDSDSDSDSDSDSDSDSDSDSDSDSD 778
QY 625 -----MLQIGESRRKRVMTLRLLLSGKADVI 650
Db 779 TSYSSAKSSVSSRSTSEASFNAILIGWRKGMRLARIGCKRVMSILRLSGKADVI 838
QY 651 KPAKRCQBEANSSGGYVQYQVNLQOQOQOQAPPPPPNPIITSPINSTLNLSLGTSGGG 710
Db 839 KGFAK-----YNEQW----- 849
QY 711 VGVGGINGFNPNTNTNTNTTGTSPSPQOQOQHGITWKSPFIPDARPRADIALYLD 770
Db 850 -----EASQSEIARYLGD 863
QY 771 IQDHITMQLLAYEKIFSRSHSNLYLAQLQVESFNNSNNKITEMFSKITLIGTMLVPLNL 830
Db 864 IQDHITMVSLSLHVKLLSRSHSNLYLAQLINIDMTKVNDMDNDVLGKITLIGTIVLPMNV 923
QY 831 VTGLFCXNVVRPCEGNTLQWPGVIGVLPFIILIGSFIQAQ 871
Db 924 ITGLMGNVIVPGYRDSLTWFIIGVLFMCHLACSAIYNTK 964

RESULT 8
ABP05970
ID ABP05970 standard; protein; 228 AA.
XX
AC ABP05970;
XX
XX 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:11922.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Shinketsu RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN21722.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,

hyperproliferative disorders and autoimmune disorders.
XX
XX Disclosure; SEQ ID NO 11922; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 228 AA;
Query Match 17.0%; Score 821; DB 5; Length 228;
Best Local Similarity 59.6%; Pred. No. 3.8e-54;
Matches 170; Conservative 23; Mismatches 32; Indels 60; Gaps 5;
QY 510 TREKVELFKSYFVFCVTFPEADKESDYLEPINVIYVPHDGLTFHFSPISHPANVRR 569
Db 1 TREKVELFKSYFVFCVTFPEADKESDYLEPINVIYVPHDGLTFHFSPISHPANVRR 60
QY 570 VRLQADVVDSADMLCYALIDEITDGFAPVHIGIEYADAEIVAFVTFARDTFSSMLQRI 629
Db 61 VRLQADVVDSADMLCYALIDEITDGFAPVHIGIEYADAEIVAFVTFARDTFSSMLQRI 120
QY 630 GSRKRVMTLRLLLSGKADVIKMFARQCEANSSGGYQOQ---YHLOQOQOQAPPPPP 686
Db 121 GSRKRVMTLRLLLSGKADVIKMFARQCEANSSGGYQOQ---YHLOQOQOQAPPPPP 173
QY 687 NPIITSPINSTLNLSLGTSGGGVGVGINFNPNTGNTNTNTNTTGTSPSPQOQOQ 746
Db 174 -----NVNQ-----NNSN-----NQI 184
QY 747 GITNKSFPIDPRADIALYLDIQDHITMQLLAYEKIFSR 791
Db 185 SILSN-SYMQSTQPRGDIYLDIQDHITMQLLAYEKIFSR 228
RESULT 9
ABP32160
ID ABP32160 standard; protein; 154 AA.
XX
AC ABP32160;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human glycoprotein-like ORF1133 protein, SEQ ID NO:2266.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; hematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;
WPI; 2001-656860/75.
N-PSDB; ABLQ4474.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

Disclosure; SEQ ID NO 7905; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABU01840-ABL16175) and the encoded proteins (AB57737-AB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1198 AA:

Query Match 4.9%; Score 234.5; DB 4; Length 1198;
Best Local Similarity 19.7%; Pred.No. 3e-08;
Matches 189; Conservative 114; Mismatches 291; Indels 349; Gaps 41;

57 SEKQQHHQBITSNAKLPLTKSGSS-----IKKS----- 87
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
37 SHQQQQQQQLMHPHKKDMLAAGSPMLPFYSHLQLOOKDATATTGPAAAAAAVEAATT 96
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
88 -----NLTDKDRITWPMISLGDDT---INSCHKRNYYN----SSLRKDFYLKN 131
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
97 SANADNFSSLIQTIDASQLGDISGLCDRFPVASFPHSNNSWTLMGTATAATTYYTN 156
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
132 TDGNTNNHTHLAIPIPIPT-----PIIVNAKSRKSGLEWLPPLIKKT 179
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
157 NNNNTNNNNNNNNVAKTPVNGSVIIESVTMPFSANILPPTHRSANEICIDPALLOKN 216
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
180 IGRNNSNNFENDLVSP-----MTKNKTNDSEDTN---TSYTAN--HNKLGIG----- 222
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
217 PQNPNGNN--SSIIVPEVHQKPLEVNSVTSSTFNLSITTAQLDLDFEQVGKDGGHI 275
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
223 ---ATTLVGVGT-----TATATATAAGR--RPRSSI----- 251
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
276 STTTTGTGSGSASGSGSGSGSGSIARTGATGATPTTYSNTANTPRTRSLHSIEELA 335
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
252 -----DSEADSHARSQSFEEDVCYPWGVDHIRVN-----GIDFD 287
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
336 ASSCAPAASPNSHTSSASTTTPQQQQQQHHQSHGNISGLSSDDSMSEDFGLEID 395
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
288 ETDETFIREEREAYLKQMTAKILRIDEFQ-----NLSKNYTSG-----ASH 332
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
396 DINGEDQQATQVFQIFSSINIPSPKOLEQQQQPPSPALAGGSSNNNAAGSNNNSAGN 455
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
333 PYHHHSNNNKNN-----GDGGGSSMAALKYT----- 360
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
456 NTSSSSNNNNNNNDNDARVLTKFBEYNAVTYLLAQGGCGSGNQOHHHSNHGHNQ 515
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
361 -----PKNILKTLRSRPFETHNSSSSSEIYEELKTKQPYP 396
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
516 QQQQQQQQQQQHQQQQQEHVQQQQQQQNLANNA--NQF-----NSSSYTYINFDQVIPT 569
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
397 KYVDQLSLTSTSTSTSGSGSQVKFGGARISDGINGS-----LPDRFLPHSESE 448
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
570 GYQD---TTSSHHSQSGGGGGG---GGGNLLNGSGGSAGGGGYMLLP----- 611
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

DE	Lactococcus lactis protein yqfG.
XX	
KW	Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX	
OS	Lactococcus lactis; IL1403.
XX	
PN	PR2807446-A1.
XX	
PD	12-OCT-2001.
XX	
PP	11-APR-2000; 2000FR-00004630.
XX	
PR	11-APR-2000; 2000FR-00004630.
XX	
PA	(INRG) INRA INST NAT RECH AGRONOMIQUE.
PA	
PI	Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX	
DR	WPI; 2002-043418/06.
XX	
PT	New nucleotide sequence useful in the identification or Lactococcus
PT	lactis and related species.
XX	
PS	Claim 6; SEQ ID NO 1665; 2504pp; French.
XX	
CC	The present invention is related to a Lactococcus lactis nucleotide
CC	sequence (AB90521) and related proteins (AB95300-AB95621). The nucleic
CC	acid sequence is useful in the detection and/or amplification of nucleic
CC	acid sequence, particularly to identify Lactococcus lactis or related
CC	species. The proteins of the invention are useful for the biosynthesis or
CC	biodegradation of a composition of interest. The invention helps research
CC	in lactic bacteria, particularly useful in the production of yogurt and
CC	cheese. Note: The sequence data for this patent is based on equivalent
CC	patent WO2001/7334 (published 18-OCT-2001) which is available in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC	standardise OS field)
XX	
SQ	Sequence 1072 AA;
XX	
SQ	Sequence 1072 AA;
XX	

Query Match	3.9%	Score	186.5	DB	5	Length	1072		
Best Local Similarity	20.8%	Pred. No.	0.00012						
Matches	193	Conservative	136	Mismatches	363	Indels	235	Gaps	45

QY	1	MDSSESYQ--NSTNOPIRSDVLDHNRQLTWN--DCAISDSEDELELKSELESEVVKSE	58
DB	305	VASSSEYQDAMSASLPISSEASSYTDNTLNSLSSDSSISSQTE-----NSQ	352
QY	59	KQQQHQQEITSDMAKPLTKRGSGSIKKKSNLTDKDRITTPKMSLGGDDTINSCHKNTNY	118
DB	353	SCASSTAELSYDSENSLSSLNQINSN--SEK--SNQSSLG---SSMSSNESEHSN	406
QY	119	MSSLRKDFYKNDTDDNSTNNHTLAIPPIPIPTIITNANKSRKKSQLENPLPKKK	178
DB	407	SSNI-----NETNNSSEI-----TNILPSPNPTESN-----VSDQ	437
QY	179	TIGRNSNPFNDLVSPWKKKTKYDSEDTWYS-----TTHNKKLGIG-----	222
DB	438	TSSSEASTNSNSISLSPNISSTDSSESATNSISDFSNVAEYANNNSLASVNNSSSVLSST	497
QY	223	--ATTGLV-----GVTGTTATATATAAGRPSSRISDSSEADSHASRSQTEEDVCPFWVG	276
DB	498	STANDLGINQSGSDNUTKDSSEITSGAFLSNQTSEASTNSNSISLSPNISSTSVL	557
QY	277	DIRVRVNGIDPDEIDFIR-----ERBEAYLQMIKAKNILRIDEP--QNLKNN---T	325
DB	558	ES--TTSSSNFNSNVAEYANNNSLASVNNSSSVLSSTSTADN--LEINQFGSNLTKDSSEIS	615
QY	326	TSQASRHPHYHHNSNNKNGDGGGGSSMAAKYTPKNTLKTLRSFEPTHENSSESSEI	385

354	QY	MAALKTKRNTLTKTLSPREPHENSSSSSEETVELTKQDPKAYDDOLSLNSTSVST---	411
355	QY		
356	QY		
357	QY		
358	QY		
359	QY		
360	QY		
361	QY	MAALKTFPNLAKTLNSREFPHENSSSSSEETVELATKQPKYKIDQSLSTSSSTSG	420
362	QY		
363	QY		
364	QY		
365	QY		
366	QY		
367	QY		
368	QY		
369	QY		
370	QY		
371	QY		
372	QY	SGSGSQVKGKARI SDG INGS L PDRFS LPHSLSSETIHADIPSLVSPQSVRDLPFN	471
373	QY		
374	QY		
375	QY		
376	QY		
377	QY		
378	QY		
379	QY		
380	QY		
381	QY		
382	QY		
383	QY		
384	QY		
385	QY		
386	QY		
387	QY		
388	QY		
389	QY		
390	QY		
391	QY		
392	QY		
393	QY		
394	QY		
395	QY		
396	QY		
397	QY		
398	QY		
399	QY		
400	QY		
401	QY		
402	QY		
403	QY		
404	QY		
405	QY		
406	QY		
407	QY		
408	QY		
409	QY		
410	QY		
411	QY		
412	QY	SGSGSQVKGKARI SDG INGS L PDRFS LPHSLSSETIHADIPSLVSPQSVRDLPFN	480
413	QY		
414	QY		
415	QY		
416	QY		
417	QY		
418	QY		
419	QY		
420	QY		
421	QY	SGSGSQVKGKARI SDG INGS L PDRFS LPHSLSSETIHADIPSLVSPQSVRDLPFN	480
422	QY		
423	QY		
424	QY		
425	QY		
426	QY		
427	QY		
428	QY		
429	QY		
430	QY		
431	QY		
432	QY		
433	QY		
434	QY		
435	QY		
436	QY		
437	QY		
438	QY		
439	QY		
440	QY		
441	QY		
442	QY		
443	QY		
444	QY		
445	QY		
446	QY		
447	QY		
448	QY		
449	QY		
450	QY		
451	QY		
452	QY		
453	QY		
454	QY		
455	QY		
456	QY		
457	QY		
458	QY		
459	QY		
460	QY		
461	QY		
462	QY		
463	QY		
464	QY		
465	QY		
466	QY		
467	QY		
468	QY		
469	QY		
470	QY		
471	QY		
472	QY	GEFTWMLDCTCPTDSEKKMLAKAFGHPHJLTADIEDTMCQETREKVELPKSYTFCVCHTTEAD	531
473	QY		
474	QY		
475	QY		
476	QY		
477	QY		
478	QY		
479	QY		

453 VKPQGVSDLPFRNGEGTWMVLCVCPDAAEKMKAAKAFGLHPLTAEDLRHQETREKVELFR 513

519 SYVYFCWTFTEADKESDYLRPIINVYIVVPHDQILTFHESPISSHANVRRVROLRDYVD 578

513 NYTVFCWTFTEGQSESDYLPIPIVYVPRGVLSFPHSPILPPIANVRRVROLRDYVD 572

579 VSADMICVYALIDETDGPAPVTHGIEVEADAEIDAPVPTARDTDFSSMLQRTIGESRRVMT 638

573 VSADMICVYALIDDTDGPAPVTVGEVEADAEIDSVPAQVGVFTWVIGRTIGESRRVMT 632

639 LMRLLLSKADVIKPKAKRCQEEA-----NSSSGCY-GRQYNLQOQQQOQAPP 683

633 LMRLLLSKADVIKPKAKRCQDEAANNOINIQAPHNNKNGMGNYTSQTSQNASQSHTHSGS 692

684 PPNPIITSPINSTLMNLNSLGTSGGSGVGSGIFGPNPPTGNTNTNTNTNTTSGSPSPQOO 743

693 YTHGYNNEPVPQCYNNNS-----NPGQYNNNSNPGQYNNNSNPGQYNNPNPQGY 743

744 QQ-----HGITNKSF-----PIPD-ARRPADIALYLGDIQDHIIITFQHLA 784

744 NOGGPGVTFPGPGGVSNMIDPGAFYMKVQTPDRQTPRADIALYLGDIQDHVVYVTFQFLVA 803

785 YKIFSRSHSNYLAQVQSPNSNNKITMFSKITLIGTLMVLPNLVTLGFMQVRRVYKGE 844

804 YKIFSRSHSNYLAQVQSPNSNYKVTAMLSNTVMTIGVTFPPPLNVTLGFMVNVYVQO 863

845 GGTN--LGMFFGIVGVLPITIGSFIPAQWMLKLNNSLEGONNGNRPFNHSSRRSIRS 902

864 GDDHPLGLMFFGIVGVCAIVAVVAVSARILVRLVTRYDEEN-----TKSKS 914

903 LGLKKH--CGNKSIIISFPNKYE 922

915 YGLTKYTRTKAAKSIISFPNKYD 936

RESULT 4

QSFV15_CANGA PRELININARY; PRT; 803 AA.

AC QSFV15; 19-JUL-2004, integrated into UniProtKB/TrEMBL.

CD 19-JUL-2004, sequence version 1.

DT 07-FEB-2006, entry version 12.

DE Similar to sp|Q08269|Saccharomyces cerevisiae YOL130w divalent cation transporter.

DE Ordered accession names: CAGLOE016178;

GN Candida glabrata (yeast) (torulopsis glabrata).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.

NCBI TaxID=3478;

RP NCITaxID=SEQUENCE [LARGE SCALE GENOMIC DNA].

RP STRAIN=ATCC 2001 / CBS 138;

RV PubMed=15229592; DOI=10.1038/nature02579;

RV La jont B, Sherman D, Fischer G, Durrens P, Casaregola S,

RA La fontaine I, de Montigny J, Mark C, Neuveglise C, Talle E.,

RA Goffard N, Pranchin L, Aigle M, Anchoard V, Babour A, Barbe V.,

RA Barnay S, Blanchin S., Beckerich J.-M., Beyne E, Bleykasten C,

RA Boisrame A., Boyer E., Catellico C., Confanioli P., de Daruvar A.,

RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,

RA Hantraye P., Hennequin C., Jeuniaux N., Joyet P., Kachouri R.,

RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,

RA Nicaud J.-M., Nikolaki M., Ostas S., Ozier-Kalogeropoulos O.,

RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,

RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,

RA Zeniou-Meyer M., Zivanovic V., Bolotin-Fukuhara M., Thierry A.,

RA Bouchier C., Saudron B., Scarpelli C., Gaillardin C., Weissenbach J.,

RA Wincker P., Couclet J.-L.,

RT 'Genome evolution in yeasts.';

RL Nature 430:35-44(2004).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

CC ENBL; CR380951; CAG58678.1; -; Genomic_DNA.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0046873; F:metal ion transporter activity; IEA.

DR GO; GO:0030001; F:metal ion transport; IEA.

DR InterPro; IPR002523; Wg2+transpCoRa.

DR Pfam; PF01544; CoRa.1.

KW Complete proteome.
SQ SEQUENCE 803 AA: 89694 MW; E741B17C22B235A4 CRC64;
Query Match 34.9%; Score 1683.5; DB 2; Length 803;
Best Local Similarity 43.0%; Pred. No. 2.9e-87;
Matches 411; Conservative 92; Mismatches 215; Indels 241; Gaps 31;
QV 38 SDEDE-----LEKSELESEVVKSEKQ-----QHQEITSNAKPLTRKSGSSIKKSN 88
DQ 3 SDESEPHSPQSHTSVAESKMSVMTDHEMLYDHRQHS-----VPPRLQDQSPGSSN 56
QV 89 LTKD-----DRITNP-----MSLGGDDTINSCHKNNYNNMSSLRKDFY-LKNDYTDNNTNN 139
DQ 57 LNARHMLDRANEDPFGMAIEGADS-----DCEPLGKDHKLASTALONSNAR 105
QV 140 HTHLAIPIT---PIPIPTIIT-NANKSRKSOLENPLPLKKTKTIGRNNNFENDLVSP 195
DQ 106 PSRLAHTVHOEVQOPATITDNNSELKNYSALSAPP-----TKTRQSNFTSTDKKP 160
QV 196 MYKMTINDSIDTSTYTAHMKLGIGATTLGVGTGTATATATAAAGRPSRSSID---252
DQ 161 L-KRTMSRHNSVNSAVSSNRLNL-----AAVSRSISHKSVDDDN 199
QV 253 -----SEADSHASRSQETEDVCFPMVGD-HIRVNGIDFIDEIDFIREEREAY 301
DQ 200 LKKRITRASFSNSQASRESQETEDVCFPMPPQLHSRVNGIDFDELEFPAEQVNE---256
QV 302 LQKOMIAKNILRIDFQNLKNNYTSGASRHPYHHNNKNGGDDGGSSN-----354
DQ 257 -QKKLC-----EISLSAKSG-----SNGIQASGSSSSANSVMS 289
QV 355 -AALKYTPKMLKTKLS--RFEPTHESSSSSEIYE-----LKTQOPPKYVDQ 401
DQ 290 EAALYTPPKH-LRKDLSPPKSESLSNNSPDSNQPGVFGNNKKEGEDP---YEEK 345
QV 402 LSVTSTSTSGSGGVKFGGARSIDGINGSLPDRFSFHSSEETIADIPSLVSP 461
DQ 346 PQAYTQTPENQFVGSGFK-----APDRFSFSESEETIATDILANLVTP 391
QV 462 GQSVRLDFRNGEETWMLDCPTDSEMKMLAKAFGIHPLTAEDIRMOETREKVELFKSVY 521
DQ 392 GQSFDFLFRGGEPTWMLDCSPTDDEMRCLAKAFGIHPLTAEDIRMOETREKVELFKSVY 451
QV 522 FVCFTTPEADKESDYLEPINVIVVYVHDGILTHPSPISHPANRRVRRLQRLVDVSA 581
DQ 452 FVCFTTPEADKESDYLEPINVIVVYVHDGILTHPSPISHPANRRVRRLQRLVDVNS 511
QV 582 DMICVALIDITGDFAPVTHGIEYEADEAVFTARDTDFSMQLQIGESRRKVTMLR 641
DQ 512 DMICVALIDITGDFAPVTHGIEYEADEAVFTARDTDFSMQLQIGESRRKVTMLR 571
QV 642 LLSGKADVIMFAKRCQBEANSQGYQRCQYNLQOQQOQAPPPPPNPIITSPINSLN 701
DQ 572 LLSGKADVIMFAKRCQBEANSQGYQRCQYNLQOQQOQAPPPPPNPIITSPINSLN 604
QV 702 SLGTSTGGVGGVGINFGNPTGNNTNTNTTSGSPPOOQOQHGTNKSFPIDARPR 761
DQ 605 -----ANLTSSDPNVKYNQKLN-----TPQR 628
QV 762 ADIALYLDIQHIIITFONLLAYEKIFSRSHSNLYLAQLQVESFNSNKKITEMFSKITLI 821
DQ 629 GDIALYLDIQHIIITFONLLAYEKIFSRSHSNLYLAQLQVESFNSNKKITEMFSKITLI 688
QV 822 GTMLVPLNLVTLGCHAVRVPEGGTGLWFFGIVGLIPIIGSFIPAFQWMLKK-----876
DQ 689 GTVLVPLNLVTLGCHAVRVPEGGTGLWFFGIVGLIPIIGSFIPAFQWMLKK-----876
QV 877 -LNNSE-----GQNGNRPPIFNHSSRRSIRSLGKKHGG--NKSIIISPNK 920
DQ 748 TLNEAASGAKSILSSPIPRGKNENDPKY-----SGRSKYGGPNSIASLPSK 798
QV 921 Y 921
DQ 799 Y 799
RESULT 5
ALRL_YEAST
ID ALRL_YEAST STANDARD; PRT: 859 AA.

Q08269; Q08211;
01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
01-NOV-1996, sequence version 1.
07-MAR-2006, entry version 41.
Magnesium transporter ALR1 (Aluminum resistance protein 1).
Name=ALR1; OrderedLocusNames=Y01130W;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Pungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RP STRAIN=S288c / FV1679;
RC MEDLINE=97051588; PubMed=8896265;
RX DOI=10.1002/(SICI)1097-0061(199609)12:10B<1013::AID-YE980>3.3.CO;2-X;
RA Casamayor A., Khalid H., Balcells L., Aldea M., Casas C., Herrero E.,
Arino J.;
*Sequence analysis of a 13.4 kbp fragment from the left arm of
chromosome XV reveals a malate dehydrogenase gene, a putative Ser/Thr
protein kinase, the ribosomal L25 gene and four new open reading
frames.
RT Yeast 12:1013-1020(1996).
[2]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=S288c / FV1679;
RC MEDLINE=97313270; PubMed=9169874;
RX Dujon B., Albermann K., Aldea M., Alexandraki D., Ansorge W.,
Arino J., Benes V., Bohn C., Bolotin-Fukuhara M., Bordonne R.,
Boyer J., Camasses A., Casamayor A., Casas C., Cheret G.,
Cziepluch C., Daiguan-Fornier B., Dang V.-D., de Haan M., Delius H.,
Durand P., Fairhead C., Feldmann H., Gaillon L., Galisson F.,
Gambo P.-J., Gancedo C., Goffeau A., Goulding S.E., Grivell L.A.,
Habbib B., Hand N.J., Hani J., Hattenhorst U., Hebling U.,
Hernando Y., Herrero E., Heumann K., Hiesel R., Hilger F., Hofmann B.,
Katsoulou C., Kordes E., Lafuente M.J., Landt O., Louis E.J.,
Maarse A.C., Madania A., Mannhaupt G., Marck C., Martin R.P.,
Kewes H.-W., Michaux G., Paces V., Parle-McDermott A.G., Pearson B.M.,
Perrin A., Pettersson B., Poch O., Pohl T.M., Poirey R.,
Portetelle D., Fajol A., Purnelle B., Ramezani Rad M., Rechmann S.,
Schwager C., Schweizer M., Sor F., Sterky F., Tarassov I.A.,
Thodou C., Tettelin H., Thierry A., Tobiasch E., Tzermia M.,
Uhlen M., Unselid M., Valens M., Vandenbol M., Vetter I., Vitek C.,
Voet M., Volckaert G., Voss H., Wambutt R., Wedler H., Wiemann S.,
Winsor B., Wolfe K.H., Zollner A., Zumstein E., Kleine K.;
*The nucleotide sequence of Saccharomyces cerevisiae chromosome XV.
RT Nature 387:98-102(1997).
[3]
NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RP STRAIN=S288c;
RC Marsischky G., Rolfs A., Richardson A., Kane M., Baqui M., Taycher E.,
Hu Y., Vannberg F., Weger J., Kramer J., Moreira D., Kelley P.,
Zuo D., Raphael J., Hogle C., Jepson D., Williamson J., Camargo A.,
Gonzaga L., Vasconcelos A.T., Simpson A., Kolodner R., Harlow E.,
LaBaer J.;
*Creation of the YLEX clone resource: cloning of Saccharomyces
cerevisiae ORFs in the Gateway recombinational cloning system.
RT Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
[4]
FUNCTION.
RP PubMed=9430719; DOI=10.1074/jbc.273.3.1727;
RC MacDiarmid C.W., Gardner R.C.;
*Overexpression of the Saccharomyces cerevisiae magnesium transport
system confers resistance to aluminum ion.
RT J. Biol. Chem. 273:1727-1732(1998).
CC -!- FUNCTION: Plasma membrane magnesium transporter.
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein
(Potential).
CC -!- SIMILARITY: Belongs to the corA family.
Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
EMBL; U41293; AAC49462.1; -; Genomic DNA.
EMBL; Z74872; CAAG9150.1; -; Genomic DNA.
EMBL; Z74871; CAAG9149.1; -; Genomic DNA.
EMBL; AY692762; AAT92781.1; -; Genomic DNA.
PIR; S66827; S66827.
GermOnline; 143552; -.

[illegible]

Db 603 IN-IA:LAQ-----SGHDP:LSARLGSNI-----CHHH-----628

Qv 754 PIPDAPRADIALYIGDIQDHITFQNLAYEKIFSRSHSNLYLAQLOVESFNSNKKITE 813

Db 629 -VORTQPRADIALYIGDIQDHITFQNLAYEKIFSRSHSNLYLAQLOVESFNSNKKITE 687

Qv 814 MFSKTTIGTLMVPLVNLVTLGFMNVVRPGEGLNGLWPGFVGVGLFIILIGSFFAQW 873

Db 688 MGLKVTMIGTLMVPLVNLVTLGFMNVVRPGEGLNGLWPGFVGVGLFIILIGSFFAQW 747

Qv 874 LKK-----LNNSEGGNNGNRPINHSRRSIRSLGKKHGGNKSIIFFPKY 921

Db 748 ISKNSPRSLNDTV---MSQTRSLINAFARNENDQC---RPEGNKSIVISCPKY 795

RESULT 8

QvCP63_KLUALA PRELIMINARY; PRT; 778 AA.

AC 16-AUG-2004, integrated into UniProtKB/TrEMBL.

DT 16-AUG-2004, sequence version 1.

DT 07-FEB-2006, entry version 11.

DE Kluyveromyces lactis strain NRRL Y-1140 chromosome E of strain NRRL Y-1140 of Kluyveromyces lactis.

GN OrderedLocusNames=KLLA0E07249g;

OS Kluyveromyces lactis (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

OX NCBI_TaxID=28985;

FN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37.

RX PubMed=15229592; DOI=10.1038/nature02579;

RA Lafontaine I., Sherman D., Fischer G., Durrens P., Casaregola S., Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C., Boissame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi R., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Nicaud J.-M., Nikolaki M., Oztas S., Ozier-Kalogreopoulos O., Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A., Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.-L.;

RT *Genome evolution in yeasts.*

RL Nature 430:35-44(2004).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms> Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; CR382125; CAG93363.1; -; Genomic DNA.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0046873; F:metal ion transporter activity; IEA.

DR GO; GO:0030001; P:metal ion transport; IEA.

DR InterPro; IPR002523; Mgf+-transptCorA.

DR Pfam; PF01544; CorA; 1.

KW Complete proteome.

SQ SEQUENCE 778 AA; 86350 MW; F68CF7FB27BA1924 CRC64;

Query Match 33.3%; Score 1603; DE 2; Length 778;

Best Local Similarity 39.9%; Pred. No. 1.1e-82;

Matches 379; Conservative 118; Mismatches 223; Indels 230; Gaps 25;

Qv 31 ITNDCAISDSEDEL-----ELKSESEVVKSEKQOHHQIITSDNAKPLRKS 79

Db 1 MSDSKSDSGEHSKRTTSAHSLSKLTETKLY--DHROOH--DYVSHNPGPAAPAA 56

Qv 80 GSGIKKSNLTOKDIRTNPM-----SLSGGDDDTINSCHKNRYNMSLSLKKDFYLKNDTD 133

Db 57 AA-----BNPQGNLGSYSNLGVVSGSGREGAAGLGVSGNFGDSD 102

Qv 134 DNSVNNHTLAIPD-----IIPPTPIITANKSRKSQENLPLPKKTKIGRNNSNF 188

Db 103 D-----LSLNPFGSNVTVIPG---SNSSK-----LGLKLVGSHSAHY 139

Qv 189 -----ENDLVSPMTKTKTNDSEDTNVTSTTANMKLGIGATTGLVGYGTTT 233

Db 140 SKSYSDGSPHRRSGHSLGILIKKSRSS---SNTSIKGNPL-----TRTVSSKSG 189

Qv 234 ATATAAAGRAPSSSIDSEADSHASRSQETEDVCFPM-VGDHVRVNGIDFDEIDEF 292

Db 190 SSKSRRLRRRLTRNASIDSD-NSQSHLSQETEDVCFPIQKREHTRINGIDFDELEEF 248

Qv 293 IREEREAYLQKQMIKNIURLIDFQNLKNNNTSGASRHPYHHHSNNKNNKNGDGG-- 350

Db 249 TAQEK-----AINNS-----HLYSHPAVTTQTTPGQHSKP 279

Qv 351 -----GSSMAALKYTPKNILKLTLSRFEETHENSSESEIEYELTKQOP-PKYDDQ 401

Db 280 PSTSDSSSSAAALKYTPR-----PEAVAQGISFGNNKVPEDYDEDDI 324

Qv 402 LSLTSSSTSGSGQVKGFGARISDGLPRPSLFLSHSESEETIHADPISLVSP 461

Db 325 KKMNSNTSTV-----PPIPBQTPFDANRPAYFSSDVSETLHAPDLSLVKP 373

Qv 462 GQSVRLDFRNGEETWLDCTCPTDSEMKMLAKAPGIIHPLTAEDIRMOETREKVELFKSY 521

Db 374 YQSPQLFNDGAPTWLDCSCPTEDEKRLTKTFTGIIHPLTSEDIRMOETREKVELFKNY 433

Qv 522 FVCFHTFEADKESDYLEPINVYVVFHDGILTFHPSPISHPANVRVRVRLQDYDVSA 581

Db 434 FICFHTFENDSESDLEPINVYVVFANGILTFHFNPSHCANVRVRVRLQDYDVNS 493

Qv 582 DMLCYALIDITDGFAPVHIGIEYADAIEDAVTARTDFFSMLQIGESRRKVTMLMR 641

Db 494 DMICYALVDDITDGFAPVQIOSIEYADAIEDSVFMTDMNFVMLQIGESRRKVTMLMR 553

Qv 642 LLSGKADVITKFAKRCQSEANSSGYSQRYNLIQOQQOQOQAPPPNPPIITSPINSLN 701

Db 554 LLSGKADVITKFAKRCQDDINGIGPALTSQVNLANLQFKMP----- 595

Qv 702 SLGTSTGGGVGGVGNFQNPNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 761

Db 596 -----DSAPSTEKROQ----- 608

Qv 762 ADIALYIGDIQDHITFQNLAYEKIFSRSHSNLYLAQLOVESFNSNKKITEMFKITLI 821

Db 609 ADIANYLIGDIQDHITFQNLAYEKIFSRSHSNLYLAQLOVESFNSNKKYTEMLKQVTMI 668

Qv 822 GTMLVPLNLVTLGFMNVVRPGEGLNGLWPGFVGVGLFIILIGSFFAQWMLKLN 879

Db 669 GTMLVPLNLVTLGFMNVVRPGEGLNGLWPGFVGVGLFIILIGSFFAQWMLKLN 728

Qv 880 ST---EQQNNGRNRPINHSRRSIRSLGKKHGGNKSIIISFP-----NKYE 922

Db 729 PTTLNEAATSGTKSIINFKRSRLKNPQNPQTKSNKSNLSFSTNMKGRYE 778

RESULT 9

QvCAS6_YARLI PRELIMINARY; PRT; 605 AA.

AC Q6CAS6;

DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.

DT 16-AUG-2004, sequence version 1.

DT 07-FEB-2006, entry version 12.

DE Yarrowia lipolytica chromosome D of strain CLIB122 of Yarrowia lipolytica.

GN OrderedLocusNames=YALI0D00319g;

OS Yarrowia lipolytica (Candida lipolytica).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Dipodascaceae; Yarrowia.

OX NCBI_TaxID=4952;

FN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=CLIB 122 / E 150;

RX PubMed=15229592; DOI=10.1038/nature02579;

RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C., Boissame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi R., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,

Matches 310; Conservative 100; Mismatches 247; Indels 220; Gaps 22;

QY 41 EDELEL-KSELESEVVKSEKQOHHQBIT-----SDNAK 73
 DB 91 ESEIELHESVMPQPKKKRRRRKASRKIPATOSSADMDLHVCPTGSCSDSKK 150
 QY 74 PLTRKS--GSGIK--KKSNUITDKDINTNPMISLGGDDTINSQKRNYN-----MSSLRK 124
 DB 151 POSNKKHGRVRKHSKSTLEVPDGPIDIKALSAMGS--SSQHASR--YNEGRGFSFDLTGT 208
 QY 125 DYPL-KUNTDSNTNNTHLAIPPIPTPIITNANKSRKQSLENLPLPKKTKTGRN 183
 DB 209 HYHLSKRSKSSSDSLVSSMSIKNDSNLSLSSSSSDVSGSDENLP--DKTLYLV 266
 QY 184 NSNNFENDLVPMTKXTNDSIEDINTSTTANMKLGIGATLGV-GTGTTATATATAA 242
 DB 267 EDPSPVHPRPSATKSCSAADCPTTTPKPPYQSDTDLTELPKTAQFSDFTTVPQ 326
 QY 243 GRPSSRSI-----DSEADSHASRSQTEEDVCFPMVGDHVRNGIDIDEIDFIREERE 298
 DB 327 RUKSTHESIADNEDREVDSQDENTRVVEEDVCFPMQESHSYKNGIDFDELNTAE-- 384
 QY 299 EAYLQKQIAKNILRIDEFQNLKNNITSGASRH-----PYHHSNNKKNNGGDDGGS 353
 DB 385 ----LQKQ-----RQNTDHFRRSQVSTCKPFPFHWNDLSHPDND----- 420
 QY 354 MAALKYTPKNILKTLRPEFTHENSSEIEYELKTKQPPKYDDQLSLTSSTSTSG 413
 DB 421 -----PSSSL-----HSNNAEAAEVLRRSS--YSGGRRSVTSMDS-- 456
 QY 414 SSGGVKFGGARISDINGSGSLPDRFSLPHSESEETHAPDIPSLVSPG-QSVRDLPFRG 472
 DB 457 -----PYRFSFFSSNQNETHASVLSLDLDDGATSFSLFCPE 494
 QY 473 EETWMLDCTCPTDSEMKLAKAFGIHPLTAEDIRMOETREKVELFKSYVFCVHTFADK 532
 DB 495 KGVWMLDCLDPTDIEMVLSKAFSIHPLTTEDIRVQAREKVELFGSYVFCVFRSFDQP 554
 QY 533 ESEYELPINVIVVFDGILTFHSPISHPANVRVRQLRDYVVSADMLCYALIDEI 592
 DB 555 ELANYLEPLMYIVVFRGLLTFHSSATHPASVRRARQLRDYVHVSDMLCYALIDDI 614
 QY 593 TDGPAVHIGIEYEADAIEDAVFTARTDPSMQLORIGESRKVMTMLLSGKADYIKM 652
 DB 615 TDAVPLIRGIEETEIEDSVLGRSESDSLRIGECRKTKMGHFRLLYKADYIKM 674
 QY 653 FAKRQBEANSSGYQRYQVNI-QOQOQOQAPPPPPNPIITSPINSLNLSLGTSTGGGVG 712
 DB 675 LAKRCNKWTIA-----PTG----- 686
 QY 713 VGGINFGPNPTGNNTNTNTNTTGTSPSPPOQOQOQHGTNKSFPIDARPRADIALYLDIO 772
 DB 687 -----PTG-----EIGLYLDIO 699
 QY 773 DRIITMFQNLAYEKIFRSNSNYLAQVFSNNKITEMFSKITLIGTMLVPLNLT 832
 DB 700 DILVTMTSNLSQFEKILSRTHSNYLAQVFSNNKITEMFSKITLIGTMLVPLNLT 759
 QY 833 GLFGNVRVPGEGTNLGMFPGVGLI-FIIIGSFI 868
 DB 760 GLFGNVRVPGSDTNLNLWPFGLGLVLLGSIAIGWII 796

RESULT 11
 Q7LWD8 SCHPO PRELIMINARY; PRT: 499 AA.
 AC Q7LWD8:
 DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
 DT 07-FEB-2006, entry version 8.
 DE Probable membrane protein YOL130w (Fragment).
 GN Name=pi066;
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.

STRAIN=972 h-;
 MEDLINE=20089027; PubMed=10620777;
 DOI=10.1007/s1097-0061(2000)011516:1<71::AID-YEA505>3.0.CO;2-5;
 Machida M., Yamazaki S., Kunihiro S., Tanaka T., Kushiida N., Jinno K.,
 Haikawa Y., Yamazaki J., Yamamoto S., Sekine M., Oguchi A., Nagai Y.,
 Sakai M., Aoki K., Ogura K., Kudo Y., Kikuchi H., Zhang M.Q.,
 Yanagida M.;
 *A 38 kb segment containing the cdc2 gene from the left arm of fission
 yeast chromosome II: sequence analysis and characterization of the
 genomic DNA and cDNAs encoded on the segment.*;
 Yeast 16:71-80(2000).
 (2)
 NUCLEOTIDE SEQUENCE.
 STRAIN=972 h-;
 Director-General of Biotechnology Center, Kushiida N., Machida M.;
 Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
 Distributed under the Creative Commons Attribution-NoDerivs license
 EMBL; AB004539; BAA21448.1; -; Genomic_DNA.
 GO; GO:0016020; C:membrane; IEA.
 GO; GO:0046873; F:metal ion transporter activity; IEA.
 GO; GO:0030001; P:metal ion transport; IEA.
 InterPro; IPR002523; Mg2+-transpCora.
 Pfam; PF01544; CorA; 1.
 FT NON_TER 1
 SEQUENCE 499 AA; 56728 MW; BGC3715EB9B6D2C4C CRC64;
 Query Match 23.7%; Score 1141.5; DB 2; Length 499;
 Best Local Similarity 40.9%; Pred. No. 1.2e-56;
 Matches 267; Conservative 60; Mismatches 143; Indels 183; Gaps 13;
 QY 232 TTATA-----TATAAGRRSRSSI-----DSEADSHASRSQTEEDVCFPMVGDHVRN 282
 DB 7 TSTAQFSDFTTVPQPKSTHSSADNEDREVDSQDENTRVVEEDVCFPMQESHSYK 66
 QY 283 GDIDPDEDFIREREAYLQKQIAKNILRIDEFQNLKNNITSGASRH-----PYHHH 337
 DB 67 GDIDPDELNPABEE-----LQKQ-----RQNTDHFRRSQVSTCKPFP 105
 QY 338 SNNKKNNGGDDGSGMAALKYTPKNILKTLRPEFTHENSSEIEYELKTKQPPYK 397
 DB 106 WNDLSHPDND-----PSSSL-----HSNNAEAAEVLRRSS--YY 139
 QY 398 YDQLSLTSSTSTSGSGSQVKFGARISDINGSGSLPDRFSLPHSESEETHAPDIPS 457
 DB 140 SGRRSVTSMSD-----PYRFSFFSSNQNETHASVLSLD 174
 QY 458 LVSPG-QSVRDLFRNGBETWMLDCTCPTDSEMKLAKAFGIHPLTAEDIRMOETREKVEL 516
 DB 175 LDDGATSFSLFCPEKGVWMLDCLDPTDIEMVLSKAFSIHPLTTEDIRVQAREKVEL 234
 QY 517 FKSYYVFCVHTPEADKSESDYLEPINVIIVVFDGILTFHSPISHPANVRVRQLRDY 576
 DB 235 FGSYYVFCVFRSFEQDPLANYLEPLNMYIVVFRGLLTFHSSATHPASVRRARQLRDY 294
 QY 577 VVVSADMLCYALIDEITDGPAPVHIGIEYEADAIEDAVFTARTDPSMQLORIGESRKV 636
 DB 295 VIVSDMLCYALIDITDAFVPLIRGIEETEIEDSVLGRSESDSLRIGECRKKT 354
 QY 637 MTLARLLSGADYIKMFARKQBEANSSGYQRYQVNI-QOQOQOQAPPPPPNPIITSPINS 696
 DB 355 MGFPELLYKADYIKLAKRCNKWTIA-----PTG----- 382
 QY 697 TLNLSLGTSTGGVGGVGINFGPNPTGNNTNTNTTGTSPSPPOQOQOQHGTNKSFP 756
 DB 383 -----PTG----- 385
 QY 757 DARPRADIALYLDIOQHITMTFQNLAYEKIFRSNSNYLAQVFSNNKITEMFS 816
 DB 386 -----EIGLYLDIOHVLVTMTSNLSQFEKILSRTHSNYLAQVFSNNKITEMFS 439
 QY 817 KITLIGTMLVPLNLTGLFGNVRVPGEGTNLGMFPGVGLI-FIIIGSFI 868
 DB 440 KITLIGTMLVPLNLTGLFGNVRVPGSDTNLNLWPFGLGLVLLGSIAIGWII 492

RESULT 12

RESULT 13

Query Match 22.88: Score 1099: DB 2: Length 663.

Matches 291: Conservative 96: Mismatches 245: Indels 260: Gaps 21:

4 SESYONSTNOPI PRSDEVLPDHRNOTTNDCA TSDSEDELEIKSELESEVVKSEKNOOH 63

7 SEKYMCLTRFSTPVE-----LDDHRPOL-----DSPPRTEATADI-SI.SRONTAQHAY 54

64 HOEITSDNAKPLTRKSGSSIKKKSNLTPKDRITNPMSLSCGDDTINSGHKNRYNMSSLR 123

db 55 HOETPO-----RPDLLSIODALREAGLSRDFEOAILD 87

124 KDFYKD-----NTDDNSTNNH-----THLAIPPIPIPTPIITNANKSRRK--SO 167

db 88 DDRSGKDINTLGRRFSDPNNGNVRHGRTWSRTHQEL-----ANMSRESSPSA 134

168 LENPPLIKKKTIGRNNNFENDLVSPMTKMKTNDESDITNTSTANHMKLGIGATTLG 227

Db 135 RSSPP-----NSVEAPADPPRRRERANTLESHPDLEA-----ILQRTVS 175

228 VGTGTTATATAAAGRRPSSIDSEADSHSRSSQTEEDVCFPVGDHVRVNGIDFD 287

db 176 GGTHTPRTFNSASA-IRPQPGDIQLEPN-----DESCVPTYEQGRIPVIDYE 223

288 EIDFIREEREAYLQKQMIAKNILRIDFQNLKNNNTSGASRHP--YHHHSNNNKKN 346

224 ELEFVA-----LSRQMKPSTSR-----KQSLSSQSRGPRVFYDLRPLRKSD 267

DR	GO: CO-0030001; P-metal ion transport; IEA.
DR	InterPro; IPR001356; Homeobox.
DR	InterPro; IPR002523; Mga2_TransactCorA.
DR	Pfam; PF01544; CorA_1.
DR	PROSITE; PS00027; HOMEBOX_1; UNKNOWN_1.
KW	Complete proteome.
SQ	SEQUENCE 460 AA; 52281 MW; 5F04F600DFD76400 CRC64;
 Query Match 21.7%; Score 1044; DB 2; Length 460; Best Local Similarity 42.5%; Pred. No. 3; 8e-51; Matches 202; Conservative 76; Mismatches 93; Indels 104; Gaps 3;	
QY	437 DRFLPFSSEETIADIPDISLVSPQSVRDLPFRNGEETWLDCCTCFPEKMKLAKAFG 496
DB	77 DHEAFPSTECETIRANEFSEIFDKCAKELFDHKGTYMLDCVDPIDAEKNLLAKAFG 136
QY	497 IHPLTAEDIRMQRTEKVELFKSYFYFCFTFEADKESDEYLEPINVIYVPFHGIILTFH 556
DB	137 IHPLTVEDIQHREAREKVEFRNYFYFAFHAQEVDSDPFMEVPPYLCVFPPEGVISPH 196
QY	557 FSPISHPANVRVRQLRVDVVSADMCLVALIDITDGPAFVINGIEVEDADEAVFT 616
DB	197 YSVPFHCAIVKKRIQLKHVTVPDMICVAVIDITDSFAPVIREIQEGSIVEQTVD 256
QY	617 ARDTDFSSMLQRTIGSRKRYTLMRLSGKAADVIMKFAKRCQEEANSSGYQRYNQVLQ 676
DB	257 ARDDGFNMLRRIGAKAKYLSMELLSGKAADVGMFTKRVSE ----- 299
QY	677 QQQQAAPPNPPIITSPTNLNLSLTGTGGVGGVGNFGPNFGNTWTNTNTVTS 736
DB	300 ----- 299
QY	737 FSPFOQQOQHGTNRKSPFIPARPRADIALYLGDITODHIITWFQNLLAYEKIFRSNSHY 796
DB	300 -----GLSDHRVPGHSILSLYGLDIODHIVTYQSLLAHKILSRSHLY 342
QY	797 LAQLQVESNNKITEMFKSITLIGTMLPVNLVTLSPGMNVRPVCGGT-----NLGMF 852
DB	343 LSQLQVSDIARTHTVDTLGKITVITGLIFPMNFTVGLFAMRVIRPGESTDENFGNMF 402
QY	853 FGTVGLVFIIPIIGSFPAQMWKLKLNISIQGNQGNRPINFNHSRRSIRSLGLXK 907
DB	403 FGLVGLVICIIVATLVARVKLAHVKKIKHN-----FNSASDNSIRMSKR 451
 RESULT 14	
ID	QAPHQ2_USTMMA PRELIMINARY; PRT; 881 AA.
AC	QAPHQ2_2005, integrated into UniProtKB/T/EMBL.
DT	15-JUL-2005, sequence version 1.
DT	07-FEB-2006, entry version 4.
DT	Hypothetical protein.
GN	GSPNames=IMG0361.1.
OS	Ustilago maydis 521.
OC	Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC	Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX	NCBL_Taxid=237631;
RN	[1]
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC	STRAIN=521;
RA	Birren B.W., Nussbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA	Ait-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA	Arachchi H.M., Armbruster J.P., Bachantang P., Baldwin J.J., Barry A.,
RA	Bavut T., Blitshteyn B., Bloom T., Byre J., Boguslavskiy L.,
RA	Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA	Calvo S.E., Camarata J., Campo K.C., Chang J., Cheshatsang Y.,
RA	Citronen M., Collymore A., Considine T., Cook A., Cooke P., Corum B.,
RA	Cuomo C., David R., Dawoe T., Degray S., Dodge S., Dooley K.,
RA	Dorje P., Dorris K., Dorris L., Duffey N., Dupes A., Elkins T.,
RA	Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA	Fitzgerald M., Foley K., Gage D., Galagan J.E., Gearin G., Gnerre S.,
RA	Gitnick A., Goymette A., Graham G., Grandbois E., Gyaltsen K., Hafez N.,
RA	Hagopian D., Hagos B., Hall J.J., Hatcher B., Heller A., Higgins H.,
RA	Honan T., Horn A., Houde N., Hughes L., Humle W., Husby E., Iliev I.,
RA	Jaffe D., Jones C., Kamal M., Katat A., Kamyesselis M., Karlsson E.,
RA	Jellis C., Kleu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,
RA	Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA	Lindblad-Toh K., Liu X., Lokysang T., Lokysang Y., Lucien O.,

QY 475 -----TWMLDCTCPTDSEKMLAKAFGIHPLTAEDIRMOET 510
 DB 549 GLSATRTALPGGIRMDPDPEANTWMLDLVSLPTQEKLSRVEFGHPLTTEDILMEET 608
 QY 511 REKVELFKSYVCFHFEADKESDYLEPIVIVVPHDGLITFHFSPISHPANVRRV 570
 DB 609 REKIELFRNYLVCFRSDQDPYSPYFLEPLANNYIIVREGTSLFPHFGPPQVARRI 668
 QY 571 RLQRYVDVSAWLCVALIDEITDGFAPVHIEYEADAEDAFTARDTDFSSMLORIG 630
 DB 669 KOLKDYINVTSDWISVALIDITDAFGPLQISIEYEVDSIDELVLILKEAEOENLARIQ 728
 QY 631 ESRKRVMTLRLLSKADVIKMPAKRQOEANSSGGYQRYQNLQOQOQOQPPPPPII 690
 DB 729 TCRKRVGLRLMGNKADVVKGLAKRCNENWS----- 760
 QY 691 TSPINSLANLSLGTSGGCVGGINFGPNPTGNNTNTNTNTTGTSPSPQOQOQHGTN 750
 DB 761 ----- 760
 QY 751 KSPFPDARPRADIALYLDIQDIHITMFONLLAYEKIFSRSHSNLYLAQVESPNNK 810
 DB 761 -----VAPKSDIGLYSDIQDIHITMTQNLNHYEILSRSHSNLYLAQSIEMTDANNQ 813
 QY 811 ITEMPSKITLIGTMLVPLNLVGLFGNNVRVPEGGNTLWGFVGVGLIFIIIGSTIFA 870
 DB 814 INDVLSKLTALGTIVPMNVITGLMGNVNVFEGVENLWPAFGICSVVMVIAVSGYIAA 873
 QY 871 QWMLKK 876
 DB 874 TRYLER 879
 RESULT 15
 ID Q2UTL7 ASPOR PRELIMINARY; PRT: 756 AA.
 AC Q2UTL7;
 DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
 DT 24-JAN-2006, sequence version 1.
 DT 07-MAR-2006, entry version 3.
 DE Hg2+ and Co2+ transporters.
 GN ORFNames=AO09009000681;
 OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 NCBI_TaxID=5062;
 [1]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=RIB 40;
 RX PubMed=16372010; DOI=10.1038/nature04300;
 RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
 RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
 RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
 RA Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
 RA Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
 RA Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
 RA Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
 RA Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,
 RA Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
 RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
 RA Komori T., Koyama Y., Minetoki T., Suhanan S., Tanaka A., Isono K.,
 RA Kuhara S., Ogasawara N., Kikuchi H.;
 RT *Genome sequencing and analysis of Aspergillus oryzae.*;
 RL Nature 438:1157-1161(2005).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AP007150; BAB55098.1; -; Genomic DNA.
 CC -----
 SQ SEQUENCE 756 AA; 85231 MW; B64BC2C7D39BA528 CRC64;
 Query Match 20.5% Score 990; DB 2; Length 756;
 Best Local Similarity 31.2%; Pred. No. 8.7e-48;
 Matches 265; Conservative 104; Mismatches 234; Indels 246; Gaps 21;
 QY 4 SSSYQNSTTNPQIPRSDVLDHNRQI-----TNDCAIS--DSEDELSKSELESE 53
 DB 7 SEKNYFLSRFSTVPPE-----LDHRRFLQDPRIEATPDMTLSRQNTTQPNLANETPQRS 62

QY 54 VVKSEKQOQHOBETSDNAPLTKRSGSSIKKSNLTKDRIITNPMSLGGDDTINSCHK 113
 DB 63 LLQVQQVQDALRE--ACGFSRDFEQAI-----ADDRSVK----- 95
 QY 114 NNNYNNLSKAFYL--KMDTNDSTNNHHTLAIPPIPTPIITNANKSRKQLENL 171
 DB 96 ----DVNGLGRFSLDPTGIRQGRAFSRTHDI-----ANMSRDS----- 133
 QY 172 PPLIKKTKTGRNNSNNENDLSPMTKNTKNDSEDIWTSTYTNHMKLGICATTGLGVGTG 231
 DB 134 --VSARSTSPNSVEAF-----ADPRRRERANTLE-----SHAAPDLAILQRTVSG-QTH 181
 QY 232 TTATATATAAGR--RPSRSSIDSEADSHASRSQETBEDVCPFMVGDHVRVINGIDFDEID 290
 DB 182 PRPPTFSNASAIRPOQDIQLDS-----PEDTCAPPFEQLGRIPVITYBELE 228
 QY 291 EPIREEREAYLQKMAKNILRIDEFQNLKSNNTTSCASRHPYHHHNNKNNKNGDGG 350
 DB 229 EFV-----ALNOKTKPTITRKHSLSQSKK----- 254
 QY 351 GSGMAALKYTPKNILKLTLSRFEPTHENSSSEIYELKTKQPPYKYDQQLSLTSSTSS 410
 DB 255 -----SRVPYDLRPNACKSEAKRASSSVSSSDDIIGLDLKDSE 295
 QY 411 TSGSGSGVQVFGGARI SDGINGSGLPDRFSLPHSSEBETHADIPSLVSPGQSVRLFR 470
 DB 296 KVFADAVNEK----QVLENLRNENEPTRFQFSSSSQSTVHAAELGDLVPPGDTPRDLFQ 351
 QY 471 NGEE--TWMLDCTCPTDSEKMLAKAFGIHPLTAEDIRMOETREKVELFKSYVFCPTP 528
 DB 352 LGPEGGVWMLDVLNPTAEVGAISRAFSIHPLTTEDILTQEAEREKVELFKQYVFCPTP 411
 QY 529 -EADKESEDYLEPIVIVVPHDGLITFHFSPISHPANVRRVRLQDYVDVSAWLCYA 587
 DB 412 YQIDKTSERFMPVNFYIVFRDGLVLSFSTENPHASNVRRIKGLRDYVLSLSDWICYA 471
 QY 588 LIDEITDGFAPVHIEYEADAEDAFTARDTDFSSMLQIGESRRKRVMTLRLLSCKA 647
 DB 472 MDDIDVDSFGPIREIEVESEADLVFIARMDDFESFLPRIGGLRKKVMSLRLGGKA 531
 QY 648 DVIRKFAKRCQEEANSSGGYQRYQNLQOQOQOQAPPPPPNPIITSPINSLNLSLGTST 707
 DB 532 DVIRGFSKRCNE----- 543
 QY 708 GGGVGVGGINFGPNPTGNNTNTNTTGTSPSPQOQOQHGTNKSFPIDARPRADIALY 767
 DB 544 -----QYSVT-----PRGDIGLY 556
 QY 768 LGDIQDIHITMFONLLAYEKIFSRSHSNLYLAQVESPNNKNTKSPKITLIGTMLVP 827
 DB 557 LGDIQDIHITMFONLLAYEKIFSRSHSNLYLAQVESPNNKNTKSPKITLIGTMLVP 616
 QY 828 LNLVTGLFG 836
 DB 617 MNLICGLFG 625
 Search completed: June 13, 2006, 14:04:46
 Job time : 321 secs

OM protein - protein search, using sw model

Run on: June 13, 2006, 14:05:02 ; Search time 47 Seconds
(without alignments)
1887.486 Million cell

Title: US-10-018-105A-4
Perfect score: 4820
Sequence: 1 MSDSSYYQNSTNOPIPRS.....LGLKKHGNKSTISFPNRYE 922

Scoring table: BLOSUM62

0
:
5
J
4
J
2
3
0
-
9
:
9
4
2
0
2
3
0

[illegible]

CHARACTERISTICS OF THE STUDY GROUP

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Database : PIR_80:*

```
2: pir2:*
```

4: pir4:*

Pred. No. is the number

and is derived by an

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1680	34.9	859	2	S66827	aluminum resistance
2	1666	34.6	858	2	S56205	probable membrane
3	1173	24.3	803	2	T40036	hypothetical protein
4	949	19.7	451	2	T37833	probable membrane
5	865	17.9	969	2	T37886	hypothetical protein
6	298	6.2	387	2	S75981	magnesium/cobalt t
7	251	5.2	351	2	H72360	divalent cation tra
8	223	4.6	319	2	H69806	divalent cation tra
9	214	4.4	380	2	S77035	magnesium/cobalt t
10	212.5	4.4	354	2	E97128	magnesium and cob
11	207	4.3	380	2	AG2319	hypothetical prote
12	200	4.1	351	2	B69348	magnesium and cob
13	192	4.0	1043	2	T13733	Frz-P1 protein - f
14	186.5	3.9	1072	2	A86827	hypothetical prote
15	184.5	3.8	1138	2	T18287	protein-tyrosine k
16	180	3.7	1212	2	T13804	shs protein - frul
17	178.5	3.7	375	2	D75591	probable cation tr
18	177.5	3.7	2452	1	RN7Q2L	DNA-directed RNA p
19	177	3.7	2457	2	T18492	hypothetical prote
20	176	3.7	1585	2	T19274	1-phosphatidylinos
21	172.5	3.6	1091	2	S57112	hypothetical prote
22	172	3.5	1368	2	T16371	probable glutamat
23	171	3.5	1368	2	T16371	hypothetical prote
24	169	3.5	1326	2	D71618	hypothetical prote
25	166.5	3.5	1858	2	C87643	hypothetical prote
26	165.5	3.4	1127	2	T28317	1-phosphatidylinos
27	164.5	3.4	2271	2	T18273	Onf MSV156 hypothe
28	163.5	3.4	956	2	S30834	hypothetical prote
29	163	3.4	852	2	T06310	hypothetical prote
30	162	3.4	1570	2	T18272	1-phosphatidylinos
31	161	3.3	366	2	F70952	hypothetical prote
32	160	3.3	369	2	D87045	hypothetical prote
33	160	3.3	373	2	T45205	divalent cation tra
34	159.5	3.3	1584	2	T18276	protein-tyrosine k
35	159.5	3.3	2150	2	S716659	sensory transducti

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

36	159	3	3	1095	2	S24061	hypothetical prote
37	158	3	3	785	2	S50016	hypothetical prote
38	158	3	3	1125	3	P90598	SO2 proteinkinase
39	158	3	3	1125	3	P90598	membrane protein
40	158	3	3	1944	2	A59438	KIAA1424 protein
41	158	3	3	1070	1	S50591	myb-related protei
42	158	3	3	1132	2	S37932	hypothetical prote
43	158	3	3	3348	1	T17414	Ti6C protein - ali
44	156	3	3	968	2	T24667	hypothetical prote
45	156	3	3	1779	2	T13613	hypothetical prote
46	155	3	3	1233	2	S56271	hypothetical prote

ALIGNMENTS

RESULT 1
 566827
 Aluminum resistance protein ALR1 - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein O0522; protein YOL130w
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 12-Jul-1996 #text_change 09-Jul-2004
 C:Accession: S66827; S63440; S71978
 C:R:Arino, J.; Casamayor, A.; Gamo, P.J.; Gancedo, C.; Lafuente, M.J.; Aldea, M.; Casas, C.; Herrero, E.; Aldea, M.; Casas, C.; Herrero, E.; Arino, J.
 Submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66814
 A:Accession: S66827
 A:Molecule type: DNA
 A:Residues: 859 <ARI>
 A:Cross-references: UNIPROT:Q08269; UNIPARC:UPI00001258CC; EMBL:Z74872;
 NID:G1209710; PIDD:CAAG9150.1; PID: c251914; PID:G1420017; MIPS:YOL130w
 A:Experimental source: strain Z28RC
 A:Reference number: S63440
 A:Accession: S63440
 A:Molecule type: DNA
 A:Residues: 1-12, 'y', 14-859 <CAS>
 A:Cross-references: UNIPARC:UPI0000168A52; EMBL:U41293; NID:G1209710;
 PIDD:AAAC49462.1; PID:G1209711
 A:Experimental source: strain PY1679
 A:R:Casamayor, A.; Khalid, H.; Balcells, L.; Aldea, M.; Casas, C.; Herrero, E.; Arino, J.
 Yeast 12, 1013-1020, 1996
 A:Title: Sequence analysis of a 13.4 kbp fragment from the left arm of chromosome XV
 A:Description: Sequencing of a 13.4 kbp fragment of the left arm of chromosome XV
 reveals a malate dehydrogenase gene, a putative Ser/Thr protein kinase, the
 ribosomal L25 gene and four new open reading frames.
 A:Reference number: S71978; MUID:97051588; PMID:8896265
 A:Accession: S71978
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-12, 'y', 14-859 <CAW>
 A:Cross-references: UNIPARC:UPI0000168A52; EMBL:U41293; NID:G1209710;
 PIDD:AAAC49462.1; PID:G1209711
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1995
 C:Genetics:
 A:Gene: SGD:ALR1
 A:Gene references: SGD:S0005490
 A:Map position: 15L
 C:Keywords: transmembrane protein
 P:743-759/Domain: transmembrane #status predicted <TM1>
 P:777-793/Domain: transmembrane #status predicted <TM2>
 C:Genetics:
 Query Match 34.98; Score 1680; DB 2; Length 859;
 Best Local Similarity 41.5%; Pred. No. 9e-95;
 Matches 405; Conservative 137; Mismatches 258; Indels 176; Gaps 31;
 Y 1 MSDSSEYQNSVTTNPQIPPSDEVLDRHQITNDCAISDSLELEL---KSELESEVKS 57
 1 MSSSSS---SSSSPNLSRS-----NSLNTATVSKNTEDHTGLVDRHQPSLPRVH 49
 DB 58 EQQQHQHQTITDNAPLRTKSGSSTKKKS-----NLTKDRIITPMSLSGDDITNSGH 112
 Y 50 OPTPLKMKIATKSPISPEKQATRYNSHVGVGSPSGRMDPDEGQDSTV-AHH 108
 DB

113 KNR -----NYNSSI-----KDYKIDKTDNSTNNHT-----HLAIPIDPIPTPI 156
 Db QLPASAILTNSARPRLSAHSPHQQLYVESNI-----HTPKDVGVKDYTMSSSTA 161
 QY TWANRSRKSOLENIPLPKIKKTIIGRNNINFDLVSFMTKTKTNDSEDITNT----- 210
 Db SSGNKS-KLSASSASPIKTVR-----KSSLVSPVLPIHESKSDTHSKLAKPKK 210
 QY -----STTAHKKLGICATTLGVTGTTATA-----TATAAARRPSRSSITSEADSHAGRSQ 264
 Db RTYSTTSAHSSIN-PAVLLTKSTQSKDADDTLKERPKVRNTRASPDSOV-SQASRDSQ 268
 QY ETEEDVCFPMVGD-HIRVNGIDFDIDEIRIREREAYLQKQMLAK-NLIRIDTFQNLK 322
 Db ETEEDVCFPMVQLHTRVNGIDFDELEVAQANAE-KSQFLASLVQNEQKYSNVSQ 325
 QY -----NNTTSGASHPHYHHNNKNNKNGDCGSSMAALKYTPKNIKLLKTLGRFEFTH 376
 Db DIGFTSSTSGSS-----AALKYTPR-VSQTGKSESTN 359
 QY ENSSSBEIYELKTKQOPPKYDDQLSSTSGSSGSGQVFGG-ARISDINGSSL 435
 Db EIEHKKDEHEKTKPSLHPISFGKNKVEGENEIPSNPDPAVCYQTDFOI 414
 QY PDRSLPHSSSETTHADIPISLVSPOQSVROLFRNGEETWMLDCTCTDSEMKMLAKAP 495
 Db PNRFFCSDETHASDIPISLVSSEGQTFELFRGCEPTWMLDSCPTDDDEMCIKAP 474
 QY GIHPTADIRMQETREKVELKFSYVCFHTEADKESDYLEPINVIVVFHDGILTF 555
 Db GIHPTADIRMQETREKVELKFSYVCFHTEADKESDYLEPINVIVVFHSGVLTF 534
 QY HPSLSPHAINVRRVRLQDYVDYSADLCYALIDETGPAVTHGIEYEADAIDEAYP 615
 Db HPGFISHCANVRRVRLQDYVYVNSDMWLCYALIDITDSFAPVQISYEYEAIDEYVP 594
 QY TARDTDFSSMLQIGESRRKQVMTLRLLSGKADVKMFARKQBEANSSGSGYQOQVNLQ 675
 Db MAHMDMAFAMLRIGESRRKQVMTLRLLSGKADVKMFARKQBEANGI----- 643
 QY QOQQQAPPPHPITSPINSTLANLSGTSFGSGVGVGINFGPNPQNNVNTVNTWG 735
 Db -----GPAUTSOIN-TANQARQNA-----SHIKNNSITVPPNY 678
 QY SPSPQOQOQHGTWKSFPIDARPDIALYLDIOHIIIMFONLLAYEKIFSRSHN 795
 Db APPTSQ-----PRGIALYLDIOHIIIMFONLLAYEKIFSRSHN 720
 QY YLAQLVLESNSNKKTFEFSKITLIGTMLDLPLKVLTCGLPGNVRVREGGCTNLCWFGFI 855
 Db YLAQLVLESNSNKKTFEFSKITLIGTMLDLPLKVLTCGLPGNVRVREGGCTNLCWFGFI 779
 QY VGLUPLIIIGSIPQAKMLKLNNSI-----EQGNKRPFN-----HSSRSISLGLKK 907
 Db LGVLLLVLVGFWPLASVRIKIDRPATLNEAESGAKSVISFFLPKRRFRNDRKNIV 839
 QY HQC-NKSLISPNKY 921
 Db RAGPSNKSVASLPSPY 855

RESULT 2
 S56205
 probable membrane protein YFL050c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein R001
 C:Species: *Saccharomyces cerevisiae*
 C>Date: 02-Sep-1995 accession_revision 19-Oct-1995 #text_change 09-Jul-2004
 C:Accession: S56205; S62287
 R:Murekami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.;
 T. Sasanuma, M.; Tsuchiya, Y.; Soeda, E.; Yokoyama, K.; Yamazaki, M.; Tashiro, H.; Eki,
 T.
 submitted to the EMBL Data Library, May 1995
 A:Description: Analysis of the nucleotide sequence of chromosome VI from
Saccharomyces cerevisiae.
 A:Reference number: S56186
 A:Accession: S56205
 A:Molecule type: DNA
 A:Residues: 1-858 <MUR>

Db 719 NTLAQGVFSNNKVTENLGGKTLGTLGTLAVLTGFGANVKKPGRNG-STAWTIG 777
 Qy 855 IGVGLVLIIGSFPAQMWLLKANSI---EQGNNGRPV-----FNHSSRSI 900
 Db 778 LGVLLLVAVISWFLASVWIKIDDPATLNEAGSGAKSVISFLPRDKRPFNDKNGN 837
 Qy 901 RSGLGKHHGKNKISIFPNKY 921
 Db 838 ARVGVR---KSTVSLPSRY 854
 RESULT 3
 T40036
 hypothetical protein SPAC27B12.12c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C:Accession: T40036
 R:Wood, V.; Rajandream, M.A.; Barrall, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.
 submitted to the EMBL Data Library, December 1997
 A:Reference number: 221900
 A:Accession: T40036
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-803 <WOO>
 A:Cross-references: UNIPROT:O13657; UNIPARC:UPI000006A290; EMBL:AL021766;
 PIDN:CAA16907.1; GSPDB:GN00067; SPDB:SPBC27B12.12c
 A:Experimental source: strain 972h-; cosmid C27B12
 C:Genetics:
 A:Gene: SPDB:SPBC27B12.12c
 A:Map position: 2
 Query Match 24.3%; Score 1173; DB 2; Length 803;
 Best Local Similarity 35.3%; Pred. No. 8e-64;
 Matches 310; Conservative 100; Mismatches 247; Indels 220; Gaps 22;
 Qy 41 EDDEL-KSELESEVVKSEKQHQHQHQT-----SDNAK 73
 Db 91 ESEIELHSEVMPQKKKKRRNKASRRKPIYSSADMDTLHVCPTGSCSDSKK 150
 Qy 74 PLTRKS---GSSIK--KKSLMDKDRITWPSLISGGDDTINSKHNNYN-----MSLRK 124
 Db 151 PQSNKKRGRRVKHSFSLVPGVDPDIALKASGKS--SQHAR--YMGREGSFTLTG 208
 Qy 125 DFLY-KONTDSTNWHHTLAIDPIPTPIPTIWAKSRKKSQLENLPLIKKIKYGN 183
 Db 209 HTLKSRSKSSDLSVSGSKIKDSSLSGSSSSSSDVGSDENLP--IDTLYLSV 266
 Qy 184 NSNPENDLVSPWTKKNTDSEDTITSTANHKLIGATLGV-CTGTATATATAAA 242
 Db 267 EDPSPVHPREPSATKSCAAVDPHTPKPPYSDTDLTLPKTAQSDFTTVP 326
 Qy 243 GRSPRSII---DSRASHASSQTEEDVCPFWGDHVRWGDIDPDTDFEERE 298
 Db 327 RUKSTHSSIAADNEBVDQDENTRVVEEDVCPHQAESHVNGKDIDELDNFAEE-- 384
 Qy 299 EAYLQKMIANKILRIDERFNLKNNITSGASH-----PYHHSSNNKKNNGGSGSS 353
 Db 385 ----LQKQ-----RQNTDFRSRQVSTCKPPEPHWDLSPHDND----- 420
 Qy 354 MAALKYTPKNILKLTLSRPEPTHNSSSSEIEYELKTKQPPYKVDQLSLTSTSTSG 413
 Db 421 -----PSSSI-----HSNNAEAAEVLRS--YVSGRRSVTSMDS-- 456
 Qy 414 SGSGQVFGGARISDGGSLPDRFSLPFSESEETHAPDPLSVSPG--QSVADLPNG 472
 Db 457 -----PVAFSPSSNQETHAHSVLSLDLDGATSPSLFCPE 494
 Qy 473 EETWMLDCTPTDSEMKLAKFGHPITARDIRMQTRKVELFKSYVFCFHTYADK 532
 Db 495 KGWMLDCLDPTDENRVLKSAFSLHPTTEDIHQEARKEVLFSGYVFCFASFEQDP 554
 Qy 533 ESEYDLEPINVYVFDGILUTPHSPISHPANVRVRRLQADYVDVSDMLCYALIDI 592
 Db 555 ELANYLSPNMYVVFREGLTLPFSSATHPASVRRARQLADYVRVSDMLCYALIDI 614
 Qy 593 TDGFAPIVHIGIEYADAEDAVFTARDTDFSSMLQRIGESRRKRVYTMALLSGKADYIK 652

C.Species: Saccharomyces cerevisiae
C.Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C.Accession: S37886
R:Rasmussen, S.; von Wettstein, D.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37872
A:Accession: S37886
A:Molecule type: DNA
A:Residues: 1-969 <RAS>
A:Cross-references: UNIPROT:P35724; UNIPARC:UPI00012P2A4; ENBL:Z28064; NID:g486086;
PIDN:CAA81901.1; PID:g486087; MIPS:YKL064w
A:Experimental source: strain 5288C
C:Genetics:
A:Gene: SGD:MNR2
A:Cross-references: SGD:S0001547; MIPS:YKL064w
A:Map position: 11L

Query Match 17.9%; Score 865; DB 2; Length 969;
Best Local Similarity 24.2%; Pred. No. 7e-45;
Matches 271; Conservative 143; Mismatches 247; Indels 460; Gaps .29;

QY	11	STTNQPIPRSDEVLDHNRQITNDCAISDSEDELELKSESEVVKSE	58
----	----	---	----

Db 44 SDSRRPTQLLHDNLQHNHGQITDFDQI-DSWGMLHESDSTSNDIIKSEDPSLKGAFIDHR 102

QY 59 ---KQQHHQEITSD-NAKPLTRKSGSSIKKKSNLTDKDRITNPM5-LSGGDDTINS GH 112

Db 103 PMSQPRGQSVSSTVQQPQIMKFTSPYKKPAGLRPSDQNRSLVSDLSPSE--LESWL 160

Qy 113 KNRNYNMSSLRKDFYLKDN-TDDNSTNNHTHLAIPPIPTPIITNA-----NKSRRK 165

D_b 161 KRR---KSVHKSFDENSPDRRQSNANDVIDVD-ALMHTVNNASTGVNDNSKRRK 215

Qy 166 SQLENPLIKKKTIGRNNNSNNFENDLVSPMTKMKTNDSEDI T N T S T A N H K L G I G A T T 225

Db 216 -----KRR--GSDSSN-KN---SKSTSSDSNDEEYNS----- 244

QY 226 LGVGTGTATATATAAAGRRPSSRIDSEADSHARSSQTEEDVCFPM--VGDIRVNG 283

DB 245 -----RPSS-----LSSNNSSLDDVCLVLDDGSEVPKAW 275

QY 284 IDPDEIFIREEREAYLQKQMIAKNILRIDEFQNLKNNNTSGASRHPYHHHSNNKK 343

Db 276 PDCTVLEEPSKEETER--LRSQAI-----QDAEAFHFQYDEDEE 312

QY 344 NNGDGGSSMAALKYTPKNILKKTLSRFEETHENSSESSEIYELKTKQPPKYDDQLS 403

Db 313 D-----GTSNEDGILFSPKPIVTNIDVPELGNRRRVNETENLKNGLRKPRIAPWHLIQRP 367

```
QY      404 LTSSTSGSGSQVKGFGARISDGINGGSL-----PD--RFSLFHSES 446
```

Db 368 VL-----GSNSTKDSKSRIQSGLQDNLLVGRNIQYPPPHIISNNPEHFRFTYFRVDL 418

447 EETIHAPDIPSLVSPGQSVRDLF----- 469

Db 419 DSTVHSPTISGLLPQPGKRFQDLFVASIYQDNSAGHIKTHPNSTPTPGIKAETVSQLQGLT 478

QY 470 -----RNGEET--WWLDC TCPTDSEMM LAKAFGIHPLTAEDIRMQETREKVE 515

DB 479 AKNPSTLSSMSVANIEDVPPFWLDVSNPTTEEMKILSKAFGIHPLTTEDIFLGEVREKVE 538

QY 516 LFKSYFVCFHTFE-----ADRES----- 534

DB 339 LFRWJYLLICFRSFDI VAEKRVRRRKRQESATLDHESISRKRKSQAYGATMSNESNANN 598

533 ----- 534

DD 399 NSISNASKSAWLFESILFRANKSSANKRIINISSSSIKRKVKSEKKKMEENEKFRKKSGRH 658

QY EDILEFINVITAVEFDGLTTFHFSPISHEFANRRVRQLCRDIYDVDSADWLCYALIDE 591
303 -
QZ

000 AFNEGEZEEFMAVATYTYVRKIOGVUTYHFAFIFHFINVRKARLEKDIENVISDWIAIALIDD 718

[illegible][illegible]

000 TACUACSTUUTVIAVUXEESTYQW

Db 332 TFIAGVYNNFKEMPE--LNSRWGYITWIVMLLIAGSLYFFWRKQWL 378

RESULT 7

H72360
divalent cation transport-related protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C:Accession: H72360
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
A:Reference number: A72200; UID:99287316; PMID:10360571
A:Accession: H72360
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-351 <ARN>
A:CROSS-references: UNIPROT:Q9WZ31; UNIPARC:UPI00000D39D8; GB:AE00512; NID:G4981074; PIDN:AAD35646.1; PID:G4981077; TIGR:TW0561
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0561
C:Superfamily: divalent cation transporter, Cora type

Query Match 5.28; Score 251; DB 2; Length 351;
Best Local Similarity 21.5%; Pred. No. 5.6e-08;
Matches 90; Conservative 63; Mismatches 131; Indels 134; Gaps 13;

QY 469 FRNGSEWLDCT-CPTDSEMKAKAFGHPHTAEIRKQREKVELFKSYVVCVCHT 527
DB 53 FADSSTPTWINTGTHRTDVQVQGEFFGHPHLEIDLNHQRKVEFFENYFVLKM 112
QY 528 FEADKEEDYLPINIVYVFDHGLTFH-----FSPISHANVRVRQRDLV-DVS 580
DB 113 FTYDKNLHE-LESEQVSLITKNCVLFQEKICDVFDP-----VRIKRYNKGIIKKR 165
QY 581 ADMLCYALIDEITGCFAPVHIEYEADAIEDAFTARDTDFSSMLORIGESRRKRVMTLM 640
DB 166 ADYLLSLDALVDYVLEKIDDEIDVLESEVLERPEK-----TVQRTHLKKNLVELR 222
QY 641 RLISGKADVTKFAKRCQBEANSSGGYQRYNQLQ000000000000000000000000 700
DB 223 KTIWPLREVL-----SSLY-----RDVPEL----- 242
QY 701 NSLGTSTGGGVGVGNGFNGPNTGNNTWNTWNTTSGSPSP000000000000000000 760
DB 243 -----YFREDVDHTIQLADVTETPRDIVSGLLDVYLSV-----SNKTNENVKVLT 296
QY 761 RADIALYLGDIQDHIITMFQNLAYEKIFSRSHSNLYLAQLOVESFNKRNKITEMFSKITL 820
DB 821 IGTMLVPLNLTGLFGNVRVFGEGTNLQWFFG-----IVGLIFIIIGSFIPAQW 872
QY 297 IATFMPLETFIAGIYGNFYMPE-----LRWKWGPVVLAVNGVIAVIMVYVFKKKW 350
DB

RESULT 8

H69806
divalent cation transport protein homolog yfjQ - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 31-Dec-2004
C:Accession: H69806
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoef, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, S.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppe, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.P.; Itaya, M.; Jones, L.; Joris, B.;

Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin, S.; A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Oudega, B.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porvolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, F.; Sekiguchi, J.; Scanlon, E. A:Authors: Schlaich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; UID:98044033; PMID:9384377
A:Accession: H69806
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-319 <KUN>
A:CROSS-references: UNIPROT:Q31543; UNIPARC:UPI00000600B8; GB:299108; GB:AL009126; NID:G2633055; PIDN:CAB12629.1; PID:G2633124
A:Experimental source: strain 168
C:Genetics:
A:Gene: yfjQ
C:Superfamily: divalent cation transporter, Cora type

Query Match 4.68; Score 223; DB 2; Length 319;
Best Local Similarity 17.7%; Pred. No. 2.5e-06;
Matches 80; Conservative 77; Mismatches 150; Indels 144; Gaps 12;

QY 430 INGSGLDPRESLPHSESEETIHAQIPSLVSPQSQVRDLFRNGEETW-WLDCTCTPTDSEM 488
DB 2 INITAITTEHQLLNKPIERVQPD-----AMVYVDFYGPDETT 42
QY 489 KMLAKAFGIHPLTADIRMOETREKVELFKSYVVCVCHTPEADKESDYLEPINVYVVF 548
DB 43 ALLRDFPHFPIAIEDCFQMRPKLDHYGVYFVYHAL-----NKETLETERVDIFQ 97
QY 549 HDGIILTFHSPISHANVRVRQRDLVDVSDMLCVLALIDEITDGFAPVHIEYEAD 608
DB 98 EKFFVTVFHLHETPGIAKVRERLYASPDILKKGPHGISYIMQDLVDEYFPLVYKIEDRLN 157
QY 609 ATEDAVTARDTDFSSMLORIGESRRKVMTLARLLSGKADVKINFAKRCQBEANSSGGY 668
DB 158 EIEE-----SRPHKTYGTGLNNEVDFDLTLHLARTIIPWRL----- 196
QY 669 QRQYNLQ000 728
DB 197 ---YRI-----LSLDRV----- 205
QY 729 TWNTTSGSPSP000 788
DB 206 -----KQRETKAYFSDIYDHLKLSSEIVSNRDM 235
QY 789 FSRSHSNLYLAQLOVESFNKRNKITEMFSKITLIGTMLVPLNLTGLFGNVRVFGEGTN 848
DB 236 TSDLADSYV-----TLNS-NRMAIWMITVTIVTIPITFTIAGVYNNFDMPE----- 284
QY 849 LGW---PFGIVGVLFIIIGSFIPAQWMLK 876
DB 285 LHWKYGVFVGLMALVIGMLI---WVHK 312
RESULT 9
S77053
magnesium/cobalt transport protein sil10671 - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein sil10671
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 31-Dec-2004
C:Accession: S77053
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraiki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996

A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions
A#Reference number: S74322; NID:97061201; PMID:8905231
A#Accession: S77053
A#Status: nucleic acid sequence not shown; translation not shown
A#Molecule type: DNA
A#Residues: 1-380 <KAN>
A#Cross-references: UNIPROT:Q55968; UNIPARC:UPI00000D344C; EMBL:D64005; GB:AB001339; NID:91001779; PIDN:BA010745.1; PID:gl006592
A#Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
A#Superfamily: divalent cation transporter CorA type

[illegible]

RESULT 10
 E97128
 magnesium and cobalt transport protein CAC1852 [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 31-Dec-2004
 C:Accession: E97128
 C:Authors: R.N.OLLING, J. Breton, G. Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, H.M.; Dubois, J.; Qiu, D.; Hitti, J.; Wolf, Y.I.; Tatusov, R.L.; Sabatche, F.; Dourette-Stamm, L.; Soucaille, P.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C1C10stridium acetobutylicum.
 A:Reference number: A96900; PMID:21359325; PMID:21359325
 A:Accession: E97128
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-354 <KUR>
 A:Cross-references: UNIPROT:Q97104; UNIPARC:UPI000000CA3D; GB:AE001437; PTDN:AA79816.1; PID:g15024829; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC1852
 C:Superfamily: divalent cation transporter, CorA type
 Query Match 4.4%; Score 212.5; DB 2; Length 354;
 Best Local Similarity 18.8%; Pred. No. 1.3e-05;
 Matches 89; Conservative 77; Mismatches 159; Indels 149; Gaps 15;

QY 423 GARISDGINSGSLPDRFSLPHSESEET-THAPDIPSLVSPGVSVDL-----PENGEEIW 476
DB 7 GUSKKKGLKPGSL-----IHVGNSTGKKTEIHIDYKDDYFEKOVHSDIDCLAYDTDNNT 63
QY 477 WLDTCTCTDSE-KMKLAKAGHPLTADIRMQETREKVELFKSVYFVCPHTPEADKSE 535
DB 64 WININGLEDIDLVEKVGKPGVHSLIMEDILNTWQRKTEEFKFIYVLKWIYP--RE 120
QY 536 DYLEPINVYVWPHDGLTIFH-----FSPISHANVRRVRQLRDV--DVSADWLKY 586
DB 121 DKLWEQVSIICWENLIITFPEEGKGDVF-----GNLRIRANTKSKIRKTCGVGLTY 174
QY 587 ALDITDGFAPVHIGTEYADAEIDAVFTARDTDFSSMLQIGESRRKRVMTLRLSGK 646
DB 175 ALDIDVONFVILEKVEDKIENFEELNT-----
QY 647 ADVIKMFAKRCQEEANGSSGYQRYNMLQOQQOQOQAPPPNPPII---TSPINSTLNLS 702
DB 205 -----TATNNILNEVYMLKRLQ-----IFLKVAVMPLREIIN--- 236
QY 703 LGTSTGGGVGVGGINFGPNPNTNTNTNTTGTSPSPQOQQOQHITNKSFPIDARPA 762
DB 237 --TLQGEVETIG-----
QY 763 DIALVLDIGDIOHITMFOQLLAYEKFSRSHSNLYAQLQVSPNSNNKITFMSKITLIG 822
DB 249 DISIVFKDVTHTVQVITTEFLRDMLAGLDTYLS-----SASNKINIKKFTIIFS 301
QY 823 TMLVPLNLTAVLFGNVRVPGGGTNGLGMFFGVIGVLVFIIGSFIFAQWLKK 876
DB 302 TIFIPSLFVGVGVGNFDYMP-----LKPRVGYAVLVMISLSAFMLMYPRKK 351

RESULT 11
AG2319
hypothetical protein al14110 (imported) - Nostoc sp. (strain PCC 7120)
A:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 sequence_revision 14-Dec-2001 #ext-change 31-Dec-2004
C:Accession: AG2319
R:Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;
Iriguchi, M.; Ishikawa, A.; Kurashina, K.; Kimura, T.; Shihida, Y.; Kohara, M.;
Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimoda, S.; Sugimoto, M.;
Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing *Cyanobacterium*
Anabaena sp. strain PCC 7120
A:Reference number: AB1807; NUTD:21595285; PMID:11759840
A:Accession: AG2319
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-380 <R>
A:Cross-references: UNIPROT:Q8YPT1; UNIPARC:UPI00000CEA09; GB:BA000019;
FIDN:BA058693.1; FID:G1133245 GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetic(s)
C:Gene(s)
C:Genefamily: divalent cation transporter, CorA type
Query Watch
Best Local similarity 4.3%; Score 207; DB 2; Length 380;
Matches 85; Conservative 71; Mismatches 126; Indels 144; Gaps 18;
QY 477 WLDTCTCTDSE-KMKLAKAGHPLTADIRMQETREKVELFKSVYFVCPHTPEADKES 534
DB 74 WDVQGLGNQDILERVQVFELPELVLEDVNVNVSREKTEDYDQLLFIARVVVPERES 133
QY 535 EDYLEPINVYVWPHDGLTIFHSPISH--ANVRRVRQLRDV--DVSADWLKYALDE 591
DB 134 GFYSQVSS--LILGKHVLLSIQEEP-EHDCPAGVVRMRKEKGGTIRQADYLAVALDA 190
QY 592 ITDGFAPVHIGTEYADAEIDAVFTARDTDFSSMLQIGESRRKRVMTLRLSGK 646
DB 191 IIDGFFV--LERYGAELEDEEVTSFSSQRT-----LQKIYHVRRLEQLRRAIWQP 242
QY 647 ADVIKMFAKRCQEEANGSSGYQRYNMLQOQQOQOQAPPPNPPIITSPIINSTLNLSLGT 706
DB 243 RDAINAL-----IDRSOLISP----- 259

Db 627 -----PTPISFGYQAYPNMNIKOBIOIPIQVSLTQSPDSFSPFI-----AIALGQ 672

Qy 703 LGTSTG-----GGVGGVGGINFGNPTGNNTNTNTNTVGS 736

Db 673 VNASGTGVIATPMNAGTGGGGGLN-GPSSVGNSSNGSSNGN 716

RESULT 14

A86827

hypothetical protein yqfG [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004

C:Accession: A86827

R:Bolotin, A.; Wincker, P.; Mauer, S.; Jalllon, O.; Malarne, K.; Weissenbach, J.; Ehrlich, S.D.; Sorokin, A.

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403.

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: A86827

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1072 <STO>

A:Cross-references: UNIPROT:Q9CF64; UNIPARC:UPI00000D491; GB:AE005176; PID:g12724625; PIDN:AAK05715.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: yqfG

Query Match 3.9%; Score 186.5; DB 2; Length 1072;

Best Local Similarity 20.8%; Pred. No. 0.0024;

Matches 193; Conservative 136; Mismatches 363; Indels 235; Gaps 45;

Qy 1 MSDESYYQ-NSTNQIPRSDVLDHNRQITN-DCAISDEDELEKSELESEVVKSE 58

Db 305 VASSSEYQDANASLIPISEASSVDTNLSISLSDSSISSQTE-----NSQ 352

Qy 59 KQQHQHQTSDNAKPIRKSGSSIKKKNLTKDRITNPMLSLGGDDTINSCHKRNYYN 118

Db 353 SGASTAEISYDSENSLSSSNQINSN-SEKD--SQNSLIG-----SSMSNEESHSN 406

Qy 119 MSLKDFYKDNNTDNNTHHAIPIPIPIPIITNANKSRKRSQLENLPLIKK 178

Db 407 SNNI-----NETNNSSEI-----TNLPSPNPTESN-----VSQO 437

Qy 179 TGRNNSNFENDLVSPMTKMTKNDSDIINTS-----TTANHKLGIG----- 222

Db 438 TSEASTNSNSISLSPSSISSTDSSEATNSDPSNVAFAVANNLSLVNNSSSVLSST 497

Qy 223 ---ATTLGV-----GTCTATATAAAGRRPSSSIDSEADSHAKRSSQTEEDVCFPMVG 276

Db 498 STADNLGINSQSDNLTQDSSEISTSGAFUSNQTSSEASTNSNSISLSPNSISSTSYL 557

Qy 277 DHRVNGIDPDEIDFIR-----EEREAYLQKMIKNIIRIDEP-----QNLKNN---T 325

Db 558 ES-ATSSSNFNSVAFAVANNLSLVNNSSSVLSSTSTADN-LEINQFGSDNLTQDSSEIS 615

Qy 326 TSGASRHPYHHHSNNKNGGSGGSSMAALKYTPKNIUKKTLRFPETHESSSSEBI 385

Db 616 TSGAFL-----SSNQTSSEASSNSKSSIN-SPSLSLST-SNSEATNQNSSEAT 664

Qy 386 YELKTKQPPYKQDQLSTSTSTSGSGSQVKGFGARISDGGISLP--DRFSLP 443

Db 665 -----KVNNSSHTSSNLTNSGSDSDSDSD-SDSNLSLSPNLETQNTS 712

Qy 444 SESEETIHAPISLVSPQSQVRLDFRNGEETWMLDCTCPTDSMKHLAKAFGLHPLTAE 503

Db 713 SKPEVANNISENPKKVSNSVQB-----NSTDHEMSTNPKSSISPSIST 758

Qy 504 DIRMQETREKVELFKSYFVCPHTFEADKESDYI-----EPINVIYVPHDGIPTFPSP 559

Db 759 SSSQO-----KESQNLNMTTEGIN-----NPITFNNS 787

Qy 560 ISHPANRRVRQLRDYVD-----VSADMLCYALIDEI--TDGPAPIVHIGIEYEAIDAYF 615

Db 788 SENG-----ASLTYSNNSESESTGCLYISNEAQRNGSEISHSLP--SSNNENNS 841

Qy 616 TAROTDFSSHQRIGESRRKVMTLRLLSKADVIKRFKRCQBEANSSSGYQRYNLQ 675

Db 842 SIQS-----QAILSSKSSTKNRSSLSIINS-----TSHRQNEQNSDSDEVKSN 890

Qy 676 QOOOQAPPPPPPIITSPINSTLNLSLGTSTGGVGGVGGINFGNPTGNNTNTNTVTVG 735

Db 891 VBS-----ILQNLNISNKNKTNMSLSQKLSVIYTLFSPSKSVTKNEKNSNTVS 939

Qy 736 -----SPSPQOOQOHG-IT-----NKSF-PIPDARPRADIAL--YLADIQDHITMF 779

Db 940 BEKLIKTPQKNDSESNLQGTALDLSFNKEVETMEDSKTVPDKVLDNENGDRSQNNKT-- 997

Qy 780 QNLAYEK---IPSSHSHNYLAQVESFNNSNKKITFMSKITLIGTMLVPLNLTGFLG 836

Db 998 -STAKDNKVPFRKRSFN-----SKLSDNNILKTKVLLKKHGLISD-----NLLIGLI- 1047

Qy 837 MNVRVPGEGTNLGWFPGVGVLPFI 863

Db 1048 -----TILGACP-----IFIFIL 1060

RESULT 15

T18287

protein-tyrosine kinase (EC 2.7.1.112) - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18287

R:Adler, K.

submitted to the EMBL Data Library, July 1996

A:Reference number: Z18856

A:Accession: T18287

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1338 <ADL>

A:Cross-references: UNIPROT:Q21927; UNIPARC:UPI000007A7F6; EMBL:U64830; MID:g1468982; PID:g1468983; PIDN:AA804999.1

C:Genetics:

A:Introns: 1181/3

C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

Query Match 3.8%; Score 184.5; DB 2; Length 1338;

Best Local Similarity 17.8%; Pred. No. 0.0043;

Matches 169; Conservative 114; Mismatches 263; Indels 405; Gaps 35;

Qy 25 DHRNQITNDCAISDSDELEKSE--LESEVVKSEKQQHQHQTTS-DNAKPLTRKSGS 81

Db 212 DTQENEDTS-----GESSDESENIDEVLYVTEDISEKKEKRRLITSPSPDPHTLYSMS 267

Qy 82 SIKKSNLTKDKRITNPMLSLGGDDTINSCHKRNYSKLRKDFYKDNNTDNNHT 141

Db 268 QSLKACSLNNNNNN--SISPSSESSINNSGNINLANSQ-----NNVNSNN-- 314

Qy 142 HLAIPIDIPPIPIITNANKSRKRSQLENLPLIKKTKTGRNNSNFENDLVSPMTOKT 201

Db 315 -----TNIINN-----SNDNFISQPLFNPLSPNN 339

Qy 202 NDSEDITNTSTTANMKLGIGATTGVTGTTATATAT--AAAGRRPSSSIDSEADSH 258

Db 340 EGYELLNPTTTTSTI-----TSITTTTITNUPPALPSPSPSSSIKLSNPF 387

Qy 259 ASRSQETEE-----DVCPPMVGDHVRNGIDPDEIDFIREEREAYLQKQIAKNI 311

Db 388 GSNSTISGELANVFSSMSPPMSPNVRSLTTPGNTPI-----NC 430

Qy 312 LRIDFQNLKNTTSGARHPYHHHSNNKNGGSGGSSMAALKYTPKNIUKKTLR 371

Db 431 INTSVINSITAN--TNCINHHHHHHQHNNHHHOG----- 463

Qy 372 FPFTHENSSESSEIYELKTKQPPYKVDQLSLSTSTSGSGSQVKGFGARISDGIN 431

Db 464 ----HNNINNSGHK-----SADDTVLTSLTSSGSSSTS----- 495

Qy 432 GGSLPDRFLSHSESEETIHAPISLVSPGQSVRLDFRNGEETWMLDCTCPTDSEMKML 491

Db 496 -----SSNPHH-----PNNHOK-- 508

Qy 492 AKAFGIHPLTAEIDIRMQETREKVELFKSYFVCPHTFEADKESDYIPEINVIYVPHDG 551

Db 509 ----GLANKTLE--KLSCTRKEI-----YELIEKKEKSLIEKQN----- 540

```

QY 552 ILTFHSPISHPANRRVRQLRDYDVVSADWLCYALIDBITDGFAPVINGIEYADAIE 611
Db 541 -----LIDE-----GISENADSPF 554

QY 612 DAVPTARDYDFSSMLQRIGESRRKVTMLMELLSGKAD-----VIKMFAPR 656
Db 555 -----NLSEETQKINE---KTIIELENLITSLNSNSNWSLNGSSTTISCNPLSPR 602

QY 657 COREANSSGYQRYMLQQQQQAPPPPNPI-ITSPINSTLNLSLGT---STGGGVG 712
Db 603 SNNPSSSTSTSNLTNSLRKFSQELKIELRPLDLRAELYSSINTSPRGSASISGSGSG 662

QY 713 VGGINFQNPFGNNTNTNT-----NTGSPSPQQQQQGIGITNKSPPIPDARPRAD 763
Db 663 GGGNNCGFKTSSNSINSPIQFFENENESIDSYEKKNEEQFESLTOLI-----RE 713

QY 764 IALYLGDIQDIHTMFQNLAYEKIFSRSHSNYLAQLQV-----ESFNSNNK 810
Db 714 NOLYTKPIE-----FREIKLEKLESNSKSNINWQIEYKSTQLVLKQPKQDQSDKIEK 767

QY 811 ITEMFSKITLIGTMLVPLNLVTLGFCNNRVFEGGTNLGWPFGIVGVLFIIIGSFIPA 870
Db 768 RKQLFN-----GSNV-----SGSN-----781

QY 871 QMWLKKLNNSIEGQNNGNRPFNHSSRRSIRSLGLKXKHGONKSIISPNNKY 921
Db 782 -----NSGSSGCGNHHHCNNS-----NGSNSEVI--PSKY 811

```

Search completed: June 13, 2006, 14:09:04
Job time : 52 secs